Metabolome analysis - mass spectrometry and microbial primary metabolites

Højer-Pedersen, Jesper Juul

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Metabolome analysis – mass spectrometry and microbial primary metabolites

Jesper Juul Højer-Pedersen

June 2008

The thesis is typeset in LaTeX

Technical University of Denmark
Center for Microbial Biotechnology
Building 223, DK-2800 Kgs. Lyngby, Denmark
http://www.cmb.dtu.dk

Summary

While metabolite profiling has been carried out for decades, the scope for metabolite analysis have recently been broadened to aim at all metabolites in a living organism – also referred to as the metabolome. This is a great challenge, which requires versatile analytical technologies that are highly sensitive and specific, and to undertake this challenge mass spectrometry (MS) is among the best candidates. Along with analysis of the metabolome the research area of metabolomics has evolved. Metabolomics combines metabolite profiles, data mining and biochemistry and aims at understanding the interplay between metabolites. In this thesis, different topics have been addressed and discussed with the aim of advancing metabolomics to explore the concept in a physiological context.

The metabolome comprises a wide variety of chemical compounds that act differently upon sample preparation, and therefore sample preparation is critical for metabolome analysis. The three major steps in sample preparation for metabolite analysis are sampling, extraction and concentration. These three steps were evaluated for the yeast *Saccharomyces cerevisiae* with primary focus on analysis of a large number of metabolites by one method. The results highlighted that there were discrepancies between different methods.

To increase the throughput of cultivation, *S. cerevisiae* was grown in microtitier plates (MTPs), and the growth was found to be comparable with cultivations in shake flasks. The carbon source was either glucose, galactose or ethanol, and metabolic footprinting by mass spectrometry was used to study the influence of carbon source on the extracellular metabolites. The results showed that footprints clustered according to the carbon source.

Advances in technologies for analytical chemistry have mediated increased amounts of data generated in high resolution. One major limitation though is the digestion of data coverting the information into a format that can be interpreted in a biological context and take metabolomics beyond the principle of guilt-by-association. To analyze the data there is a general need for databases that contain metabolite specific information, which will speed up the identification of profiled

metabolites. To address the capabilities of electrospray ionization (ESI)-MS in detecting the metabolome of *S. cerevisiae*, the *in silico* metabolome of this organism was used as a template to present a theoretical metabolome. This showed that in combination with the specificity of MS up to 84% of the metabolites can be identified in a high-accuracy ESI-spectrum. A total of 66 metabolites were systematically analyzed by positive and negative ESI-MS/MS with the aim of initiating a spectral library for ESI of microbial metabolites. This systematic analysis gave insight into the ionization and fragmentation characteristics of the different metabolites. With this insight, a small study of metabolic footprinting with ESI-MS demonstrated that biological information can be extracted from footprinting spectra. Statistical analysis of the footprinting data revealed discriminating ions, which could be assigned using the *in silico* metabolome. By this approach metabolic footprinting can advance from a classification method that is used to derive biological information based on guilt-by-association, to a tool for extraction of metabolic differences, which can guide new targeted biological experiments.

Dansk sammenfatning

Analyse af metabolitter og sammenligning af profiler af disse har været udført i årtier, og inden for de senest år er der kommet fornyet fokus på denne disciplin, hvor man nu ønsker at måle samtlige metabolitter i en celle — også kaldet metabolomet. Det er en betydelig udfordring at analysere metabolomet, og det stiller store krav til analysemetoderne, der skal besidde stor følsomhed og specificitet. Der er et begrænset antal analyseteknikker til rådighed for at klare denne udfordring og heriblandt må massespektrometri anses for en af de bedre muligheder. Hånd i hånd med analysen af metabolomet er der opstået et forskningsområde kaldet metabolomics, hvor metabolitprofiler, dataanalyse og biokemi kombineres for at opnå en helhedsforståelse af metabolitternes sammenspil. I denne afhandling diskuteres forskellige emner inden for metabolomics med henblik på at bringe metabolomics et skridt videre i forståelse af, hvordan det kan bruges i en fysiologisk sammenhæng.

Metabolomet er sammensat af en lang række kemisk forskellige forbindelser og prøveforberedelse er derfor kristisk for metabolomanalyse. De tre primære trin i prøveforberedelse er prøveudtagning, ekstraktion og opkoncentrering. Disse tre trin er blevet undersøgt for gæren *Saccharomyces cerevisiae* med henblik på at måle så mange metabolitter så muligt på en gang. Forsøgene understregede at de forskellige metoder gav forskellige resultater.

Med henblik på at kunne gennemføre flere forsøg blev *S. cerevisiae* dyrket i mikrotiterplader, og væksten synes at være sammenlignelig med vækst i en rysteflaske. Som en del af forsøget blev der brugt 3 forskellige kulstofkilder: glukose, galactose og ethanol og kulturvæsken blev analyseret for ekstracellulære metabolitter ved metabolic footprinting. Data blev opnået ved massespektrometri, og dataanalysen viste, at det var muligt at skelne prøverne, efter hvilken kulstofkilde der var brugt.

Inden for analytisk kemi sker der store tekniske fremskridt, og det har bl.a. betydet at data bliver generet i stadig højere opløsning. Dette kræver en effektiv fortolkning af data, således at det bliver muligt at ekstrahere biologisk informa-

tion. Denne fortolkning kan assisteres af databaser, der indeholder referencedata, og disse databaser vil kunne effektivisere dataanalysen og identifikationen af metabolitter. For at belyse muligheden for at bruge elektrosprayioniseringmassespektrometri (ESI-MS) til detektion af hele metabolomet for *S. cerevisiae* blev der taget udgangspunkt i *in silico* metabolomet af denne organisme. Det viste sig, at med specificiteten fra massespektrometri var det muligt at skelne mellem op til 84% af metabolitterne i et massespektrum. Yderligere blev der indsamlet data fra 66 metabolitter med henblik på at lave en database med ESI-MS spektre for mikrobielle metabolitter. Data blev systematisk genereret, og med udgangspunkt i disse data har det været muligt at forbedre fortolkningen af metabolomdata opnået ved massespektrometri. Som et eksempel herpå var det muligt at udlede biologisk information på baggrund af metabolic footprinting. Normalt bruges metabolic footprinting til klassifikation, men med udgangspunkt i massespektrene blev det muligt at tage metabolic footprinting skridtet videre og udlede fysiologiske forskelle.

Preface

The present thesis concludes 3 years research started in March 2003 at Center for Microbial Biotechnology resided at the department Sysmtems Biology at Technical University of Denmark (DTU). The work has been carried out under supervision by Professor Jens Nielsen and co-supervision by Associated Professor Jørn Smedsgaard.

The thesis comprises the primary work of several distinct projects aiming at advancing metabolomics to explore the concept in a physiological context. The research has been a journey within bioanalytical chemistry spiced with microbial physiology and informatics. Mass spectrometry has had a central position in the work, which is reflected in the thesis.

Acknowledgements

Numerous people have been interacting with me during the study. Let me start by recognizing Professor Jens Nielsen as being my supervisor throughout the project. I thank him for the many good and fruitful discussions we had along the way. I appreciate his support when it was needed in the very end. Associated Professor Jørn Smedsgaard has been co-supervisor throughout the project, and I thank him for generously sharing his expertise in analytical chemistry and especially mass spectrometry.

Professor Gregory Stephanopoulos from Massachusetts Institute of Technology (MIT) is thanked for giving me the opportunity to stay at MIT for four months.

I also dedicate my thanks to Michael Edberg Hansen, Silas G. Villas-Bôas and Sandrine Mas for close collaborations and stimulating scientific discussions. I would like to thank visiting student Nicolas Moreau for his contributions to my thesis. Furthermore, I am appreciative for the many good discussion with Kristian Fog Nielsen about strategies for analytical chemistry. I am also thankful for the technical support from Kianoush K. Hansen and Hanne Jakobsen.

I have sincerely enjoyed my time at Center for Microbial Biotechnology thanks to the enjoyable people at the center. Officemates and colleagues are thanked for scientific and social moments.

Last but not least it would not have been as fun without proper econ-*omics* and therefore I am grateful to the Technical University of Denmark for providing the PhD scholarship. Federation of European Microbiological Societies (FEMS), Otto Mønsteds Fond, Augustinus Fonden, Familien Hede Nielsens Fond, and Reinholdt W. Jorck og Hustrus Fond are all acknowledged for financial support for conferences and travels.

Kgs. Lyngby, June 2008

Jesper Juul Højer-Pedersen

Contents

Publications

Peer-reviewed papers

1. S. G. Villas-Bôas, J. Højer-Pedersen, M. Åkesson, J. Smedsgaard, and J. Nielsen. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* 22: 1155–1169, 2005.

2. J. Højer-Pedersen, J. Smedsgaard, and J. Nielsen. The yeast metabolome addressed by electrospray ionization mass spectrometry: Initiation of a mass spectral library and its applications for metabolic footprinting by direct infusion mass spectrometry *Metabolomics* (in press)

3. K. B. R. M. Krogh, P. V. Harris, C. L. Olsen, K. S. Johansen, J. Højer-Pedersen, F. Tjerneld, and L. Olsson. Cloning and characterization of a GH3 β-glucosidase from *Penicillium brasilianum* including a novel method for measurement of glucose inhibition of cellobiose hydrolysis. (submitted)

Book contributions

4. J. Højer-Pedersen, J. Smedsgaard , and J. Nielsen. Elucidating the mode-of-action of compounds from metabolite profiling studies. In H. I. Bashoff, and C. E. Barry III, editors, *Progress in Drug Research*, volume 64, pages 105–129. Birkhäuser Verlag, 2007.

Poster presentations

5. J. Højer-Pedersen, J. Smedsgaard, and J. Nielsen. Analysis of amino acids by ion trap LC-MS. *Physiology of Yeast and Filamentous Fungi II*. Marts 23–28, 2004. Anglet, France.

6. J. Højer-Pedersen, J. Smedsgaard, and J. Nielsen. Mass spectrometry for metabolite fingerprinting. *Metabolic Engineering V*. September 18–23, 2004. Lake Tahoe, CA, USA.

Other publications

7. J. Højer-Pedersen, J. Smedsgaard, and J. Nielsen. Metabolomet: Et indirekte produkt af genomet. *Dansk kemi* 87(3): 18–20, 2006.

Nomenclature

Roman Letters

e	elementary charge of an electron	-1.6022×10^{-19} C
E_{acc}	energy from acceleration	
m	mass	
m/z	mass-to-charge ratio determined by mass spectrometry	
q	electrical charge	
t	time	
U	electrical potential	
V	amplitude of RF voltage	
v	velocity	
z	number of elementary charges	

Greek Letters

ω	frequency

Abbreviations

ANOVA	analysis of variance
API	atmospheric pressure ionisation
BPC	base peak chromatogram

CE	capillary electrophoresis
CID	collision-induced dissociation
DC	direct current
EI	electron impact ionisation
EIC	extracted ion chromatogram
ESI	electrospray ionisation
FDA	Fisher discriminant analysis
FWHM	full-width half maximum
GC	gas chromatography
GTP	guanosine triphosphate
LC	liquid chromatography
MCA	metabolic control analysis
MS	mass spectrometry
MTP	microtiter plate
NMR	nuclear magnetic resonance spectroscopy
PC	principal component
PCA	principal component analysis
PEG	poly(ethylene glycol), $H(OCH_2CH_2)_nOH$
Q	quadrupole
QIT	quadrupole ion trap
RF	radio frequency
SIM	selected ion mode
TIC	total ion chromatogram
TOF	time-of-flight

Chapter 1

Introduction

In the 1990'ies major advances were made in genomic research. Automated gene sequencing seeded the initiation of several sequencing programs with the objective to sequence whole genomes. In 1996 the whole genome for baker's yeast *Saccharomyces cerevisiae* was published as the first fungal genome and also the first genome of a eukaryotic organism (Goffeau et al., 1996). At present (January 2008) 708 complete genome sequences are published (http://www.genomesonline.org) and the number is rapidly increasing. Genome sequences have provided substantial biological information, which has led to comparative studies across species to decode the four-letter code in the genetic material. This has been one of the major driving forces in initiation of systems biology, since the genomic sequence (genotype) sets the theoretical boundaries for the cellular capabilities. However, based on the genome sequence alone it is not possible to predict all resulting phenotypes; although the genome provides the code for life and in other words forms the basis for phenotypic behavior. The genetic information flow in a cell starts by transcription of the genes to mRNA that is further translated to proteins (Figure 1.1). The proteins interact with present small molecules (metabolites) that react and form a metabolic network. On top of this the environment and a wide range of regulatory mechanisms play a role through interactions – everything contributing to the phenotype.

To undertake the challenge of predicting phenotypes we must take a holistic approach inspecting the whole cell or organism as a system – moving away from the reductionistic approach that has been widely applied in biology (Stephanopoulos et al., 2004). To face this challenge of global biology a variety of techniques have been developed that allow high-throughput, multi-parallel analysis of biological systems, *e.g.* hybridization arrays (Fodor et al., 1993; Schena et al., 1995; de Saizieu et al., 1998), LC-MS/MS protein analysis (Gygi et al., 1999; Washburn

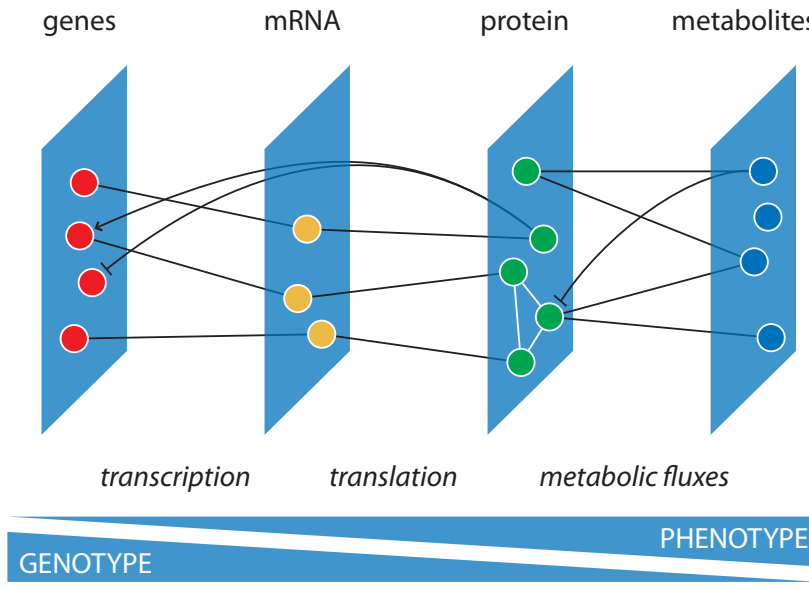

Figure 1.1 The central dogma of biology including metabolites. The genes give the potential of an organism (genotype) and provide the boundaries for the metabolic capabilities, whereas proteins and especially metabolites reflect the actual activities under a given set of conditions (phenotype) as they assemble the functional metabolic network.

et al., 2001), metabolite profiling platforms (Fiehn et al., 2000; Raamsdonk et al., 2001) and interaction assays (Fields and Song, 1989; Uetz et al., 2000). The ability to generate high-throughput data have reversed research strategies towards data-driven approaches where new hypothesis are generated based on data (Kell and Oliver, 2004). This is in contrast to hypothesis-driven research that bases the experimental design on a hypothesis.

1.1 Metabolite profiles

Metabolite profiling studies have been carried out for decades, but recently metabolite profiling has attracted increased attention as a tool in functional genomics and systems biology (Fiehn et al., 2000; Raamsdonk et al., 2001; Roessner et al., 2001; Trethewey, 2001; Weckwerth and Fiehn, 2002; Weckwerth, 2003). The term metabolome has been coined to describe the complete pool of metabolites in a cell, and the metabolome is basically the sum of all metabolites in analogy to the terms genome, transcriptome and proteome (see Table 1.1 for definitions). It

Table 1.1 Definitions used in this thesis in relation to metabolites analysis

Term	Definition
Metabolite	A low-molecular-weight chemical compound that is related to metabolism, but not directly derived from the genome and nor structurally bound in the cell
Metabolome	The complete set of metabolites associated with a cell/organism at any given time. The metabolome is subdivided into the endo- and exo-metabolome that cover the intra- and extra-cellular metabolites, respectively
Metabolite profiling	Chemical analysis of numerous metabolites
Metabolic fingerprinting	A chemical signature or snapshot of (intracellular) metabolites in crude extracts. The fingerprints are mainly used for classification and biomarker discovery
Metabolic footprinting	Metabolic footprinting is equivalent to metabolic fingerprinting, but covers analysis of extracellular metabolites. The term is primarily used in microbial metabolomics
Metabolomics	The research field focussing on comprehensive chemical analysis of the metabolome and integration of data through data mining and biochemistry (Figure 1.2)

covers the comprehensive set of all low-molecular-weight molecules (metabolites) associated with a cell at any given time. The metabolome can be divided into the endo- and the exo-metabolome that cover the metabolites inside and outside the cell, respectively. In these sub-metabolomes some metabolites are present in both the endo- and the exo-metabolome, since they are transported across the cell membrane. A major part of the metabolome (the primary metabolites in the central metabolism) is conserved across species, and this property makes it possible to develop generic analytical techniques that can be applied across species even when the genome sequence is unknown.

The central dogma in biology hierarchically links genes, transcripts and proteins, but as such the metabolites cannot be characterized as a direct product of proteins (Figure 1.1). Nevertheless, proteins catalyze the conversion of metabolites, which make metabolite profiles an imprint of physiological states. The catalytic activity of enzymes is a prerequisite for metabolic reactions, and metabolites are therefore indirectly a downstream product of gene expression (Fiehn, 2002; Schmidt, 2004; Nielsen and Oliver, 2005; Schreiber, 2005). Metabolites are the intermediates of biochemical reactions that are parts of metabolic pathways. These pathways are interconnected and hereby constitute a metabolic network

that is the functional machinery in a living cell. Metabolite concentrations are determined by the enzyme kinetics of the different pathways that produce and consume the metabolites, and hence the metabolite concentrations are indirectly influenced by gene transcription, mRNA translation, protein-DNA interactions, protein-protein interaction, posttranslational modifications *etc.*

The metabolic network of *S. cerevisiae* has recently been reconstructed *in silico* (Forster et al., 2003). The model represents the current knowledge of the major biochemical reactions present in *S. cerevisiae* and consist of approximately 1,200 reactions involving around 600 metabolites. Using this model the connectivity of the metabolic network can be inspected using graph theory, which results in a very compact graph (Nielsen, 2003; Nielsen and Oliver, 2005). The metabolic reactions are incorporated in the graph by connecting the reaction substrates to the catalyzing enzyme and further on the reaction products. This way the web of metabolic reactions are represented in a bipartite graph consisting of both metabolites and enzymes as nodes. Analysis of this graph shows that more than 70% of the metabolites are participating in more than just two metabolic reactions and thus indicating the compact and highly connected structure of the graph. Given this fact, it is evident that only a minor perturbation of the metabolic network will result in a global response in the metabolite pools. This can be seen as both an advantage and disadvantage. The advantage is that the metabolite levels are very likely to reflect a change in *e.g.* the genome, but on the other hand it is difficult to locate the reason for the change due to the wide-spread effects it might have.

Metabolite concentrations are the variables that to some extend control the metabolic fluxes. From the theory of metabolic control analysis (MCA)[1] it is known that changes in enzyme activities might have a little effect on the related fluxes, whereas they are very likely to have significant effects on the concentrations of metabolites. Teusink et al. (1998) applied MCA to theoretically investigate a metabolic network perturbation. Specifically, mutants of the 2,6-bisphosphofructokinase encoded by *PFK26* and *PFK27* in the glycolytic EMP pathway were investigated. The theoretical results of the MCA showed that steady-state metabolite concentrations of the pathway would be affected by the mutation and this was later experimentally confirmed by Raamsdonk et al. (2001). In addition to this Raamsdonk et al. (2001) introduced a functional genomic approach by ^{1}H-NMR (nuclear magnetic resonance spectroscopy, see Section 3.1.2) analysis of metabolite extracts and solely based on the NMR data it was possible to cluster the different mutants. And even analysis of extracellular metabolites using direct infusion mass spectrometry (MS) proves to allow discrimination of various

[1] MCA is the sensitivity analysis of metabolic fluxes and metabolite concentrations in metabolic pathways.

mutants (Allen et al., 2003).

Metabolite quantities provide integrative information of the many different processes operating in a living cell, and therefore they are very likely to be reporters of developmental, genetic and environmental changes, which to some degree is complementary to the genetic information (Oliver et al., 1998; Raamsdonk et al., 2001; Allen et al., 2003, 2004).

1.2 Metabolomics

The definition of metabolomics and related terms are still a matter of some debate, and to enlighten the definitions Table 1.1 contains a list of the definitions used in this thesis. Metabolomics has been defined as the quantitative analysis of the complete metabolome (Fiehn, 2002), which is equivalent to global metabolite profiling. I argue that there is more to metabolomics than just comprehensive metabolite profiles. Metabolomics should be considered as a research area rather than an analytical approach as already discussed by Villas-Boas et al. (2005). Of course comprehensive chemical analysis of the metabolite pools (metabolite profiling) is a prerequisite for metabolomics research to give global metabolic states; but along with this, data mining and biochemistry will have a major impact on the interpretation of data, which is required for metabolomics to succeed. In Figure 1.2 the three major components of metabolomics: chemical analysis, data mining and biochemistry are integrated.

Metabolomics covers an integrated approach, which includes the functional level of metabolites in a cell. Genome-scale models of metabolic networks can be constructed through combination of genome sequences, biochemistry and literature (Borodina and Nielsen, 2005; Edwards et al., 1999; Palsson, 2000), and these comprehensive models represent the scaffold for the metabolome and opens up for integration of upstream functional levels like proteins, mRNAs and genes. The genome-scale models contain a large number of metabolites and these metabolites can serve as templates driving the metabolome analysis.

A major challenge at present is to fully access the metabolome and there is still some way to go to be able to completely cover the full metabolic network. A number of analytical techniques has to come into play to measure the metabolome. MS has for many years had a central position in metabolite analysis, and MS is expected to be one of the primary technologies for analysis of the metabolome, since MS has the sensitivity and specificity required for metabolome analysis.

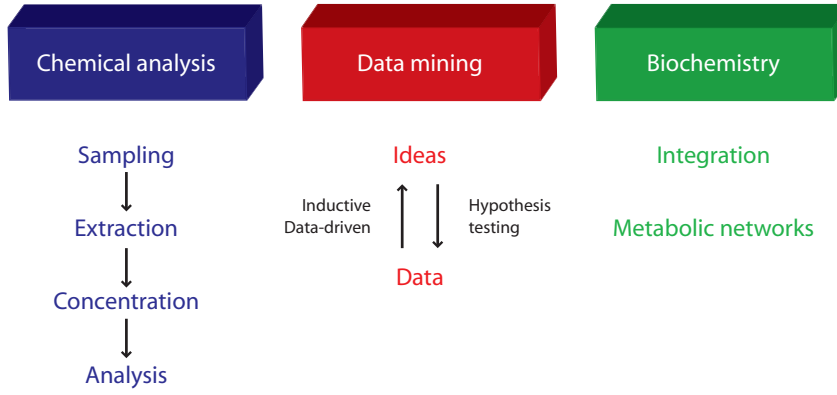

Figure 1.2 Metabolomics is comprehensive metabolite profiling in combination with data mining, which can be structured by biochemistry. Both metabolite data and biochemistry has to be combined and integrated in the data mining.

1.3 Thesis overview

The aim of the research, which is presented and discussed in this thesis, has been to explore strategies to pursue metabolites as genetic markers, since there is an increasing interest for metabolites as functional markers. The potential is large and the power of comprehensive metabolite profiling has already been shown. While the degree of insight gained from the metabolite profiles is still limited, this thesis communicates different concepts that have been investigated to improve the interpretation of metabolome data with focus on MS.

The thesis composes eight chapters, which can be read independently. After this introductory chapter, the following three chapters present background information and discuss aspects of MS, metabolite profiling and data analysis with emphasis on metabolomics. Chapter 2 provides some basics of MS in relation to metabolite profiling and is primarily devoted to technological descriptions. Then Chapter 3 proceeds by reviewing different approaches for metabolite profiling with emphasis on mass spectrometry associated techniques and a few examples of nuclear magnetic resonance spectroscopy (NMR) – another powerful technique for metabolomics. The chapter is partly adapted from Publication 4. Hereafter, Chapter 4 outlines considerations regarding data handling and analysis with specific emphasis on MS data.

A considerable part of chemical analysis of metabolites is sample preparation and Chapter 5 deals with sample preparation protocols for the yeast *S. cerevisiae*. This work results from a collaboration with former PhD student Silas G. Villas-

Bôas and the chapter corresponds to Publication 1.

Chapter 6 describes the growth of *Saccharomyces cerevisiae* in microtiter plates, which potentially can ease the handling of many samples. The results emphasize the potential of using different carbon sources. The experimental work was assisted by visiting student Nicolas Moreau.

The next chapter, Chapter 7, is a submitted manuscript (Publication 2), which is addressing the capabilities of electrospray ionization MS for detection of the yeast metabolome. Furthermore, the manuscript is an initiation of construction of a spectral library containing tandem mass spectra of primary metabolites. The metabolites are analyzed by electrospray MS in both positive and negative mode. It is a considerable task to cover the full metabolome and at present data for 66 metabolites are contained in the library. The application of the list of yeast metabolites is then used to deduce biological information for metabolic footprint of *S. cerevisiae*.

Finally, the thesis is concluded in Chapter 8 by discussion of the presented results in Chapters 5–7. Further the concept of dynamic footprinting is presented with some preliminary results arising from a collaboration with Sandrine Mas and Michael Edberg Hansen. Dynamic footprinting extends the metabolic footprinting approach by investigating the dynamics in metabolic footprinting during the time course of a batch cultivation.

References

J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.*, 21(6):692–696, 2003.

J. Allen, H. M. Davey, D. Broadhurst, J. J. Rowland, S. G. Oliver, and D. B. Kell. Discrimination of modes of action of antifungal substances by use of metabolic footprinting. *Appl. Environ. Microbiol.*, 70(10):6157–6165, 2004.

I. Borodina and J. Nielsen. From genomes to *in silico* cells via metabolic networks. *Curr. Opin. Biotechnol.*, 16:350–355, 2005.

A. de Saizieu, U. Certa, J. Warrington, C. Cray, W. Keck, and J. Mous. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nat. Biotechnol.*, 16(1):45–48, 1998.

J. S. Edwards, R. Ramakrishna, C. H. Schilling, and B. O. Palsson. *Metabolic engineering*, book chapter Metabolic flux balance analysis, pages 13–57. Marcel Dekker, 1st edition, 1999.

O. Fiehn. Metabolomics - the link between genotypes and phenotypes. *Plant Mol. Biol.*, 48(1-2):155–171, 2002.

O. Fiehn, J. Kopka, R. N. Trethewey, and L. Willmitzer. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.*, 72(15):3573–3580, 2000.

S. Fields and O. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989.

S. P. Fodor, R. P. Rava, X. C. Huang, A. C. Pease, C. P. Holmes, and C. L. Adams. Multiplexed biochemical assays with biological chips. *Nature*, 364(6437):555–556, 1993.

J. Forster, I. Famili, P. Fu, B. O. Palsson, and J. Nielsen. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.*, 13(2):244–253, 2003.

A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science*, 274(5287): 546–567, 1996.

S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, 17(10):994–999, 1999.

D. B. Kell and S. G. Oliver. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*, 26(1):99–105, 2004.

J. Nielsen. It is all about metabolic fluxes. *J. Bacteriol.*, 185(24):7031–7035, 2003.

J. Nielsen and S. G. Oliver. The next wave in metabolome analysis. *Trends Biotechnol.*, 23(11):544–546, 2005.

S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz. Systematic functional analysis of the yeast genome. *Trends Biotechnol.*, 16(9):373–378, 1998.

B. Palsson. The challenges of *in silico* biology. *Nat. Biotechnol.*, 18:1147–1150, 2000.

L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.*, 19(1):45–50, 2001.

U. Roessner, A. Luedemann, D. Brust, O. Fiehn, T. Linke, L. Willmitzer, and A. Fernie. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, 13(1):11–29, 2001.

M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467–470, 1995.

C. W. Schmidt. Metabolomics: what's happening downstream of DNA. *Environ. Health Perspect.*, 112(7):A410–A415, 2004.

S. L. Schreiber. Small molecules: the missing link in the central dogma. *Nat. Chem. Biol.*, 1(2):64–66, 2005.

G. Stephanopoulos, H. Alper, and J. Moxley. Exploiting biological complexity for strain improvement through systems biology. *Nat. Biotechnol.*, 22(10):1261–1267, 2004.

B. Teusink, F. Baganz, H. V. Westerhoff, and S. G. Oliver. Metabolic control analysis as a tool in the elucidation of the function of novel genes. In A. J. P. Brown, M. F. Tuite, and J. Prosser, editors, *Methods in Microbiology*, volume 26, pages 297–336. Academic Press, 1998.

R. N. Trethewey. Gene discovery via metabolic profiling. *Curr. Opin. Biotechnol.*, 12 (2):135–138, 2001.

P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403(6770):623–627, 2000.

S. G. Villas-Boas, S. Rasmussen, and G. A. Lane. Metabolomics or metabolite profiles? *Trends Biotechnol.*, 23(8):385–386, 2005.

M. P. Washburn, D. Wolters, and J. R. Y. 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.*, 19(3):242–247, 2001.

W. Weckwerth. Metabolomics in systems biology. *Annu. Rev. Plant Biol.*, 54:669–689, 2003.

W. Weckwerth and O. Fiehn. Can we discover novel pathways using metabolomic analysis? *Curr. Opin. Biotechnol.*, 13(2):156–160, 2002.

Chapter 2

Mass spectrometry for metabolomics

Comprehensive metabolite profiling requires sensitive and specific methods that can determine a wide range of chemical compounds varying in polarity and concentration range. Mass spectrometry (MS) fulfills these requirements to a large extent and development of biological MS has been one of the major driving forces in the development of metabolomics. MS presents a versatile technology and mass spectrometers can be used to resolve complex samples either on its own or in connection with separation techniques (Chapter 3). The mass spectrometer provides high sensitivity and at the same time structural and chemical information.

The basic function of a mass spectrometer is to determine the mass-to-charge ratio (m/z) of charged compounds, where m is the ion mass measured in daltons[1] (Da) and z is the number of elementary charges. To do this, the mass spectrometer needs an ion source, where the ions are formed; a mass analyzer in which the ions are separated by m/z; and a detector quantifying the ion abundance by measuring the ion current or count. Between the ion source and the mass analyzer there are a series of ion lenses and ion optics that focuses and guides the ion beam to the mass analyzer situated in the high vacuum envelope of the mass spectrometer. A schematic overview of the mass spectrometer is shown in Figure 2.1.

In this section, some basics on MS technologies within the field of metabolite profiling will be presented by mainly describing ionization techniques and mass analyzers.

[1]The mass unit dalton (Da) is defined to be 1/12 of the mass of one ^{12}C atom equal to 12×10^{-3} kg mol^{-1}, thus 1 Da $= 1 \times 10^{-3}$ kg mol$^{-1}/N_{\mathrm{A}} = 1.661 \times 10^{-27}$ kg, where $N_{\mathrm{A}} = 6.022 \times 10^{23}$ mol^{-1} is Avogadro's constant.

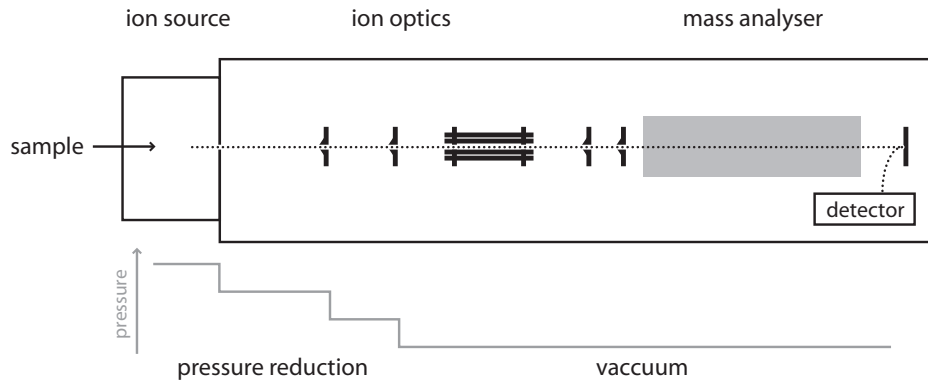

Figure 2.1 Schematic overview of a mass spectrometer. The mass spectrometer consists of an ion source where the compounds are ionized, and a series of ion lenses and ion optics that guides the ions to the mass analyzer and the detector *in vacuo*. A series of vacuum and turbo pumps maintain the vacuum.

2.1 Ionization processes

To measure the mass by MS, it requires a charged molecule, an ion, present in the gas phase. There are multiple ways to produce ions for mass spectrometric detection; however, for metabolite profiling electron impact ionization (EI) and electrospray ionization (ESI) are predominantly used when connected to gas and liquid chromatography, respectively.

2.1.1 Electron impact ionization

EI is a classical ionization technique and is used in connection with gas chromatography (GC). The EI process takes place *in vacuo* where the analyte molecule is ionized by passing a beam of energized electrons emitted from a heated filament. The electrons are accelerated to an energy of 70 eV (electron volts). When an electron hits the molecule, it may knock of another electron and hereby ionize the molecule by production of a positively charged radical ion (Equation 2.1). The radical ion ($M^{\bullet+}$) is called the molecular ion and has the same molecular mass as the molecule (except the mass of an electron[2]).

$$M + e^- \longrightarrow M^{\bullet+} + 2e^- \tag{2.1}$$

[2]The mass of an electron is $m_e = 5.486 \times 10^{-4}$ Da.

The impact with energetic electrons at 70 eV leaves some of the molecular ions with excess energy that is sufficient to mediate bond breakage and subsequent production of smaller fragment ions (Equation 2.2). The ionization energy of 70 eV is high compared to the energy of a covalent bound (typically 3–7 eV), and therefore the EI process is considered as a hard ionization process. By EI, most ions produced are single charged and therefore the m/z determined by MS conveniently corresponds the m of the (fragment) ion. However, due to the relatively high ionization energy the molecular ion is predominantly fragmented and therefore not always present in the spectrum to reveal the molecular mass.

$$\begin{array}{rcl} \mathrm{M}^{\bullet+} & \longrightarrow & \mathrm{M}_1^+ + \mathrm{M}_2^\bullet \\ & & \\ & & \mathrm{M}_3^+ + \mathrm{M}_4 \\ & \nearrow & \\ \mathrm{M}_1^+ & \longrightarrow & \mathrm{M}_5^+ + \mathrm{M}_6 \\ & \searrow & \\ & & \mathrm{M}_7^+ + \mathrm{M}_8 \\ & \vdots & \\ & \textit{etc.} & \end{array} \tag{2.2}$$

The fragmentation pattern (relative occurrence of fragment ions) will be characteristic for the molecule and comparable to a molecular fingerprint well suited for identification. The ionization energy of 70 eV is the standard energy for EI and the ionization has proven to be very robust and give reproducible fragmentation patterns over time and for different instruments. This has allowed compilation of comprehensive spectral databases/libraries enabling confident identification of compounds (Halket et al., 2005).

2.1.2 Electrospray ionization

For many years, coupling of liquid chromatography (LC) to MS was hindered by the incompatibility of liquid flows at atmospheric pressure and vacuum in the mass spectrometer. Simple injection of liquid into vacuum induces an extensive volume expansion and results in collapse of vacuum. Introduction of atmospheric pressure ionization (API) techniques in the mid 1980'ies came around these obstacles and revolutionized the mass spectrometric analysis of biomolecules, because it enabled ionization from a liquid phase and an easy coupling of LC to MS.

The most popular API technique is ESI, which is frequently used in biological mass spectrometry. In ESI, high voltage (2–5 kV) is applied to a conducting capillary through which a liquid solution is pumped and electrospray occurs from the tip of the capillary. The mechanism underlying ESI is still a matter of some debate and not fully understood, but there is a general consensus that the ESI combine three major processes: i) the production of charged droplets from solvent containing electrolytes, ii) droplet shrinkage by solvent evaporation and repeated droplet disintegration (fission) leading to highly charged droplet from which iii) gas-phase ions are produced (Kebarle and Tang, 1993).

The electrical potential at the tip of the capillary partly separates negative and positive ions in the solution by electrophoretic movement. In positive mode, positive ions are enriched at the liquid surface at the capillary exit. The over-representation of positive charge at the surface mediates repulsion and along with the pull of the electrical field, the liquid elongates into filaments – referred to as the Taylor cone. At a certain point the force of charge repulsion and attraction to the counter-electrode exceeds the surface tension and small droplet are emitted from the tip of the Taylor cone as illustrated in Figure 2.2. This produces a fine spray of droplets with a surplus of positive charge. The droplet size is dependent on the liquid composition, the flow rate, the viscosity and ionic strength in the liquid. Mostly, the droplet diameter is around 1–2 μm, when spraying from normal size capillaries around 0.1–0.2 mm in inner diameter (Kebarle, 2000; Cole, 2000; Cech and Enke, 2001). In nano-electrospray very small needles less than 10 μm in outer diameter produces droplets with diameters in the nm-range.

The droplets shrink in size due to solvent evaporation and consequently the charge density increases, which leads to Coulomb fission at 70%–80% of the Rayleigh limit[3]. The Coulomb fission produce tiny off-spring droplets from a "tail" (Taylor cone) of larger non-spherical droplets (Gomez and Tang, 1994) as illustrated in Figure 2.2. The process repeats for the smaller droplets and ultimately, highly charged and very small droplets of few nanometers in size are formed from which gas-phase ions are produced. The actual mechanism for production of gas-phase ions from tiny charged droplets is still not fully understood, but there seems to be two models describing the process, the ion evaporation model and the charge residue model. In brief, 1) the ion evaporation model proposes emission of single gas-phase ions directly from tiny charged droplets when these reach a certain diameter, and 2) the charge residue model proposes that ultimately only one charged residue is left in the droplet and hence after evaporation of the final solvent the charged residue (ion) is present in the gas phase (Cole, 2000; Kebarle and Tang, 1993; Kebarle, 2000).

[3]The Rayleigh limit is the droplet radius at which the force of electrostatic repulsion equals the surface tension force.

While the positive charges emerge from the Taylor cone at the tip of the capillary, oxidation processes releasing electrons occur at the interface between the capillary and solution. *E.g.* in aqueous solutions, the oxidation of water releases H^+ that can affect the pH and subsequently the mass spectrum through changed ionization properties.

The ESI mechanisms are less studied in negative mode and at present also less frequently used than positive mode. For simplicity only the mechanism has been described for positive ESI, but the fundamental mechanism is the same in negative mode just with opposite charges, *i.e.* in negative mode the electrophoretic movement will be reversed, so that the negative ions will accumulate on the surface of the liquid filament and reduction reactions will occur at the capillary wall.

Many factors influence electrospray and do have an impact on the spray, *e.g.*

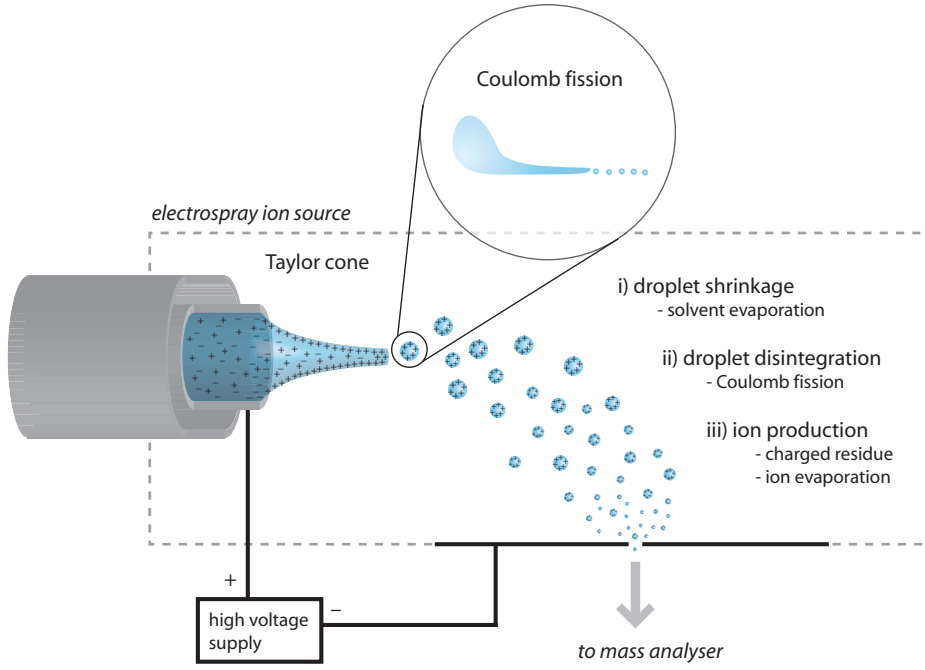

Figure 2.2 Electrospray ionization is performed by applying high voltage to a conducting capillary through which a solution is flowing. Through formation of a liquid filament called the Taylor cone the liquid emerge into fine charged droplets that ultimately yield ions in the gas phase. The figure summarizes the overall ESI process in positive mode. The circle in the top illustrates the generation of off-spring droplets from a tail of a larger, non-spherical droplet by Coulomb fission (Gomez and Tang, 1994).

the organic solvent content in the liquid, modifiers, the surface activity of the analyte, the ionic species, the ionic strength *etc.* To obtain electrospray and formation of the Taylor cone it is required to have electrolytes (ions) present in the solvent, but the ionization is matrix dependent and one molecule does not necessarily yield a unique mass spectrum, which makes ESI spectra less robust compared to EI for application of database searches for identification. *E.g.* trace amounts of sodium ions can suddenly result in increased sodium adduct formation in ESI. As a rule-of-thumb volatile acids improve the ionization in positive mode by providing excess of protons to charge the compounds. For negative ESI, Wu et al. (2004) studied the influence of different modifiers on the negative electrospray ionization and found that weak acids can have a favorable effect on the ionization efficiency.

The ESI process depends strongly on the composition of the solution, the so-called matrix effects, that can change the ionization efficiency and lead to suppression or amplification of the ionization. One major reason for these phenomena is changes in the ionic strength and surface tension in liquid that will affect the charge distribution in the droplet and on its surface caused by modifiers or co-eluting substances (Matuszewski et al., 2003; Taylor, 2005). High concentrations of buffers and especially non-volatile buffers are incompatible with ESI and will leave you with low signal, if any at all. The buffer will simply snatch the charge from the molecules of interest. Application of volatile buffers consisting of *e.g.* formic or acidic acid and ammonia is however a way to reduce the ion suppression caused by buffers (McCalley, 2003). A recommended concentration range of electrolytes for ESI is around 10^{-5}–10^{-3} M (Kebarle, 2000).

The ionization can be obtained by protonation $[M + H]^+$ or deprotonation $[M - H]^-$, but adduct formation with *e.g.* Na^+, NH_4^+ or Cl^- is common. Solvent adducts may also appear, *e.g.* $[M + CH_3CN + H]^+$, $[M + CH_3CN + Na]^+$ and $[M+CH_3OH+H]^+$, and more complicated ions like $[2M+H]^+$ and $[M-H+2Na]^+$ are also found. The latter may indicate the presence of a carboxylic acid group, where the carboxylic acid group is dissociated and binds a sodium ion apart from the whole molecule forms a sodium adduct. For larger biomolecules like peptides and proteins multiple charged ions $[M + nH]^{n+}$ are very common.

2.2 Mass analyzers

After production of the ions by ionization in the ion source the ions enter the mass analyzer. The mass-to-charge ratios of the ions are determined by guiding the ions through a mass analyzer to a detector, where a computer is used to acquire the mass spectrum by summarizing the ion statistics. There are several mass

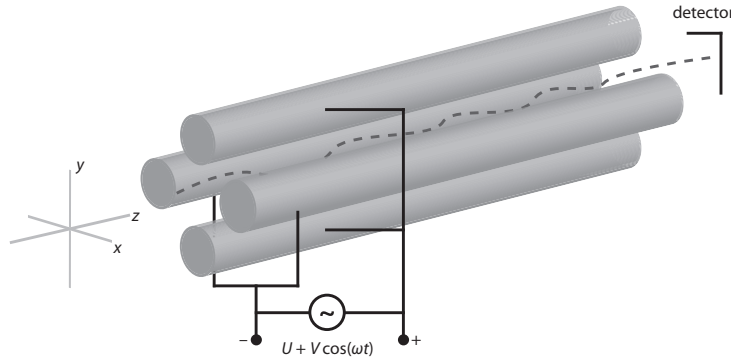

Figure 2.3 The quadrupole is constructed of four metal rods with an DC and RF voltage applied. The ion path through the quadrupole is determined by the DC and RF voltage applied to the rods. The resonating (stable) ions will follow a fluctuating trajectory through the rods of quadrupole and hereafter reach the detector as illustrated by the dashed line. Unstable ions will collide with the rods and not reach the detector.

analyzer technologies available for separation of ions prior to detection and in this section the quadrupole, the ion trap and the time-of-flight (TOF) technologies are described. A summarizing overview of the mass analyzers are given in Table 2.1 in the end of the section.

2.2.1 Quadrupole

The quadrupole is one of the technically simpler mass analyzers available and it presents a versatile and robust technology that is fairly easy to operate. The quadrupole acts as a mass filter that allows passage of ions with a set mass-to-charge ratio to reach the detector. By scanning the filter mass, spectra covering an m/z range from 2 Da/e up to about 4000 Da/e can be recorded. For a standard instrument, unit mass resolution allows mass determination within 0.3 Da, but in the recent years high-resolution quadrupole instruments have also become available.

The quadrupole is constructed of four parallel metal rods arranged as illustrated in Figure 2.3. The two sets of diametric rods are electrically connected and a radio frequency (RF) voltage $V\cos(\omega t)$ and an off-set (DC) voltage U (electrical potential) is applied across the two pairs (Figure 2.3). Before entering the quadrupole, the ions are accelerated to the same translational energy and they enter the quadrupole through a hole that determines the width of the ion beam. In the quadrupole, the ions start to fluctuate in the xy-plane (Figure 2.3), while

the movement along the z-axis is ideally unaffected by the quadrupole field and determined by the acceleration voltage applied before the quadrupole entrance. The trajectory of the ion flight path through the four rods is given by the amplitude of the oscillating RF potential and the constant DC potential; and depending on the m/z of the ion, it will either resonate and pass through the quadrupole for detection or collide with one of the rods (Steel and Henchman, 1998).

The quadrupole can be operated as mass analyzer in two modes referred to a selected ion mode (SIM) and scan mode. In SIM, the U and V are kept constant and tuned to allow transmission of a specific m/z. This results in very sensitive and specific detection of the selected ion as all the analysis time is spend to monitor this particular ion. No spectrum is recorded and thus no identification and structure elucidation is possible. Therefore, SIM is primarily used for targeted analysis or trace analysis, where maximal sensitivity is required. To record a mass spectrum, the quadrupole is scanned (scan mode) by ramping the DC and RF voltages, while maintaining the ratio between those (U/V) constant. Hereby a m/z range can be covered and combined into a mass spectrum. The larger the mass range the longer the scan time, which decreases the sensitivity because relatively less time is used on measuring the individual masses.

2.2.2 Ion trap

The ion trap is comparable to the quadrupole and can be described as a three-dimensional quadrupole that traps the ions in a potential well, formed by a three-dimensional electrical field. To be more correct the ion trap is actually named a quadrupole ion trap (QIT). In contrast to the quadrupole that continuously transmits resonating ions the QIT works by first trapping the ions in the potential well and then ejecting the ions to reach the detector. Hence, the full duty cycle of a QIT consists of ion injection, ion trapping, and finally ion ejection for detection.

The QIT is constructed of a ring electrode and two end-caps that form a cavity in the center of which the ions can be trapped (Figure 2.4). A DC and RF potential $U + V\cos(\omega t)$ is applied to the ring electrode and the end-caps are grounded (March, 1997). The potentials define the ion trajectories and in regions of stable ion trajectories, the ions will be trapped and move in a complex saddle-form fashion within the trap (Forbes et al., 1999). The maximum mass range for trapped ion is obtained by only applying the RF potential and leaving the DC potential at 0 V (March, 1997). The trapped ions can then be ejected from the trap for acquisition of a mass spectra. The ion trajectories are destabilized by increasing amplitude of the RF voltage, which will lead to ejection of ions to the detector in ascending order according to m/z. The ejection is caused by destabilization in the direction of the end-caps and the ions are ejected through

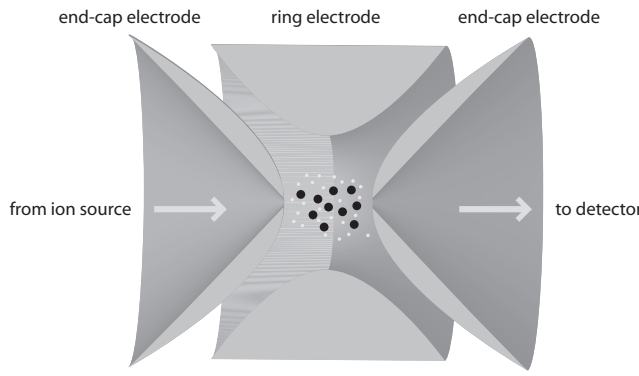

Figure 2.4 The (quadrupole) ion trap is constructed of a ring electrode and two end-caps that form a cavity in which ions can be trapped. The ions arrive from the ion source and are trapped in stable cloud in presence of helium (small white dots) to decrease the kinetic energy of the ions (black circles). By manipulation of the electrical field in the trap the ions are ejected to the detector in an ascending order according to mass-to-charge ratio.

holes in the end-caps (March, 1997).

Helium gas is present in the ion trap at around 0.13 Pa to reduce the kinetic energy and stabilizing the ions in the trap. The helium gas helps dapping the motion of ions by low-energy collisions and concentrate them in the center of the trap, which allows a faster duty cycle for trapping and ejection.

A strong feature of the QIT is its ability to perform multiple fragmentation cycles (MS^n). The ion of interest is isolated in the QIT and by increasing the kinetic energy, fragmentation is induced by collision of the ion with the helium already present in the trap. After collision-induced fragmentation, the QIT contains the daughter ions that can be ejected to the detector giving a tandem mass spectrum. This cycle can be repeated and thus enabling MS^n (Jonscher and Yates, 1997; March, 1997).

2.2.3 Time-of-flight

The TOF analyzer is far more intuitive and easy to understand than quadrupole and QIT. The ions are directed into a flight tube of a certain length and hence the flight time can be related to the m/z of the ions. There are no complicated trajectories that arise from electrical or magnetic fields constraining the ions in the flight tube. After acceleration, they follow at simple flight path at constant velocity.

Ions carrying a charge of $q = z\ e$ with z being the number of elementary

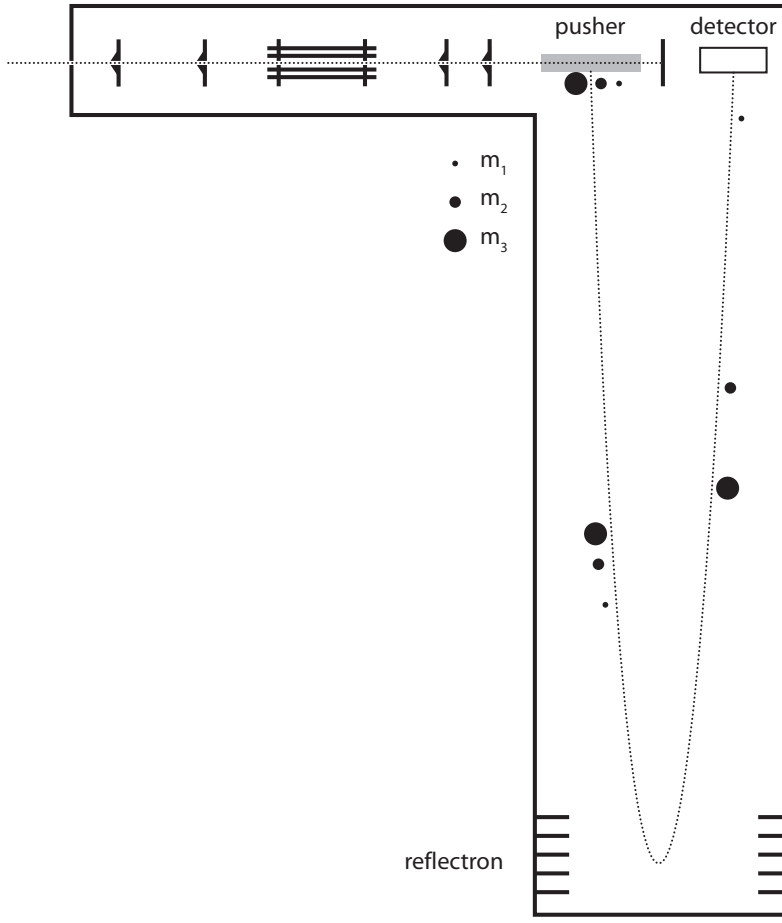

Figure 2.5 The time-of-flight (TOF) mass analyzer determines the mass-to-charge ratio of ions by measuring the time it takes the ions to fly along an evacuated flight tube. The pusher pushes the ions into the flight tube and when they reach the reflectron in the end of the flight tube they are reversed and directed to the detector. The reflectron functions as an electric mirror that reduces dispersion in energy and improves mass resolution by lengthening the flight path.

charges and e the elementary charge of an electron, are accelerated through an electric potential (U) to obtain the translational energy (E_{acc}) of

$$E_{acc} = q\ U = z\ e\ U, \tag{2.3}$$

which equals the kinetic energy relating the mass m and velocity v as

$$E_{kin} = \frac{1}{2}\ m\ v^2. \tag{2.4}$$

After acceleration, the velocity is constant *in vacuo* and it can simply be determined by measuring the time of flight (t) of the ion in an evacuated flight tube of constant length (d):

$$d = v\ t. \tag{2.5}$$

Combining Equations 2.3, 2.4 and 2.5 gives the relation between the mass-to-charge ratio (m/z) and time of flight:

$$m/z = \frac{2\ e\ U}{d^2}\ t^2, \tag{2.6}$$

where it is found that m/z is proportional with the squared flight time. Thus, there is higher spacial resolution between lighter ions compared to heavier ions.

Instruments tuning is important for optimal performance of mass analyzers. During tuning of TOF instruments, the ion optics are optimized to transmit ions from the ion source to the mass analyzer in a narrow beam – the narrower (and thereby less disperse) a beam is the higher the mass resolution. The beam enters a pusher that by an instantaneous electric pulse pushes an ion package orthogonal to the beam direction (Figure 2.5). The pusher potential determines the acceleration that ideally should give ions the same energy (E_{acc}) and hence, ions with the same m/z should obtain the same velocity and reach the detector at the same time. However, in reality the velocity varies and some equivalent ions are slower and some faster, which eventually will affect the mass resolving power. To correct for the variation in energy a reflectron can be inserted in the end of the flight tube. The reflectron acts like an electric mirror that doubles the flight path (and time) and at the same time also evens out velocity variation for equal mass ions. The reflectron reverses the ion movement and the reflection is obtained by repulsion in a series of increasing electrical fields. The faster the ion the earlier the ion reaches the reflectron and also the deeper the ion penetrates into the electrical fields and *vice versa* for the slower ions. Thus, the flight distance gets adjusted according to the actual ion energy to compensate for energetic dispersion, and consequently the ions are focussed resulting in higher mass resolution.

From each pusher impulse, all ions are pushed into the flight tube and the flight time for ions of 1000 Da/e is typically less than 50 μs, thus next push events can proceed right after all ions has reached the detector. From each push a mass spectrum can be assembled and summarizing several push events into one mass spectrum improves the ion statistics and reduces the noise level (and the required disc space on the computer). The fast push events make it possible to acquire spectra with short time intervals, which is ideal in combination with fast separation techniques. Unlike the quadrupole, there is no advantages of using SIM since all ions are pushed into the flight tube and needs to reach the detector before the next push event.

2.3 Ion identification

The power of mass spectrometry is that it produces compound specific information related to the elementary composition of the ion, and fragmentation reflects compound structures. This does not only make detection by MS highly specific, but can provide molecular information about unknown compounds (Fiehn et al., 2000; McLafferty and Turecek, 1993).

All mass spectrometers have sufficient mass resolution to separate isotopes (at least in the mass range of metabolites) and therefore they do not measure the average molecular weight of ions, but rather the monoisotopic masses and distinguish the different isotopes. For this purpose Table A.1 in Appendix A lists the accurate mass and isotopes of the typical atoms in biomolecules.

Multiple charged ions can be identified from the isotope masses. If there is less than one mass unit between the isotopes, then the ion is multiple charged, *e.g.* if an ion is double charged the isotopes are separated by half of a mass unit.

Compounds containing an odd number of nitrogen atoms have a resulting odd molecular mass. This is referred to as the nitrogen rule and can be helpful to indicate the presence of nitrogen. Furthermore, high-resolution MS gives accurate mass and makes it possible to determine candidates for the molecular composition by searching the combinatorial space of molecular compositions in a narrow mass range determined by the mass accuracy. This is indeed valuable information in relation to identification of unknown components. The existence of different isotopes of the same atom gives rise to isotope patterns that can support the composition of a given ion in the mass spectrum. Thus, not only the mass, but also the relative abundances of the isotope masses can be used to confirm the actual ion composition (Kind and Fiehn, 2006; McLafferty and Turecek, 1993). Especially, chloride, bromide and sulfur have characteristic isotope patterns.

As already mentioned in Section 2.1.2, ESI produces different adducts. Ta-

Table 2.1 Summarizing overview of common mass analyzer technologies in relation to metabolite profiling

MS technology	Description
Quadrupole (Q)	The Q act as a mass filter that only allows transmission of one mass-to-charge ratio to the detector at any time. The Q is constructed by four metal rods on which an oscillating and a constant potential is applied to control the transmission of ions. The Q can be operated in single ion mode (SIM) for monitoring a specific mass or in scan mode to acquire a mass spectrum. The Q is typically an analyser with unit mass resolution and the accuracy of mass is within 0.3 Da.
Quadrupole ion trap (QIT)	The QIT operates by trapping the ions and sequentially ejecting them to the detector where they are quantified. Additionally, the QIT enables fragmentation experiments that are useful for determining structures or for increased specificity. Theoretically there are no limitations on the number of fragmentations one can perform, thus MS^n is possible. Similarly to the quadrupole, the QIT normally operates at unit mass resolution and accuracy within 0.2 Da.
Time-of-flight (TOF)	Acceleration of ions to the same kinetic energy and subsequent measurement of time of flight in a evacuated flight tube of constant length makes it possible to determine the ion mass through the expression for kinetic energy $E_{kin} = \frac{1}{2}\ m\ v^2$. The TOF technology enables high-resolution data and mass accuracies below 5 ppm can be obtained.
Hyphenated technologies	Combination of the above mentioned technologies expands the MS capabilities. Triple quadrupole detectors are suited for tandem MS and allow high sensitivity for trace analysis. The Q-TOF configuration enables tandem MS with determination of accurate masses of fragments. These are just two examples of hyphenated technologies.

Table 2.2 Typical adduct and fragment ions produced from electrospray ionization (ESI) of primary metabolites

	Positive ESI		Negative ESI	
	Ion composition	Mass change	Ion composition	Mass change
Adducts	$[M+H]^+$	+ 1	$[M-H]^-$	− 1
	$[M+Li]^+$	+ 7	$[M+Cl]^-$	+ 35[a]
	$[M+NH_4]^+$	+ 18	$[M+CHOO]^-$	+ 45
	$[M+Na]^+$	+ 23	$[M+CH_3COO]^-$	+ 59
	$[M+K]^+$	+ 39	$[M+HSO_4]^-$	+ 97[a]
	$[M-H+2Na]^+$	+ 45	$[M+H_2PO_4]^-$	+ 97
	$[M-2H+3Na]^+$	+ 67		
	$[M-3H+4Na]^+$	+ 89		
Fragments	$[M-NH_3+H]^+$	− 16	$[M-H_2O-H]^-$	− 19
	$[M-H_2O+H]^+$	− 17	$[M-H_3PO_4-H]^-$	− 99
	$[M-CO+H]^+$	− 27		
	$[M-CO_2+H]^+$	− 43		
Multimers	$[2M+H]^+$	+ M + 1	$[2M-H]^-$	+ M − 1
	$[2M+Na]^+$	+ M + 23		

[a] Adduct ions have characteristic isotope pattern according to the adduct molecule

ble 2.2 summarizes the most frequent adducts and lists some of the common fragments that can be used for ion identification and help to derive the molecular mass.

2.4 Mass spectral data

For processing of data this section will give a brief overview of formats of mass spectral data and considerations regarding comparison of multiple spectra. Mass spectral data come in two forms: continuum or centroid. Continuum data correspond to what is acquired by the mass spectrometer and centroid data are a processed and compressed form that contains the centroid masses as illustrated in Figure 2.6. The centroid data contain the primary information from the continuum data and efficiently save disc space on the computer.

For automatic comparison of two mass spectra binning is used to align data. In order to bin the data a mass grid is defined and then the ions from the mass spectrum are assigned to the corresponding mass bins (Figure 2.6). This way mass spectrum is converted from a continuous scale to discrete bins which can easily be aligned and organized into a matrix for several spectra. If more than one ion falls

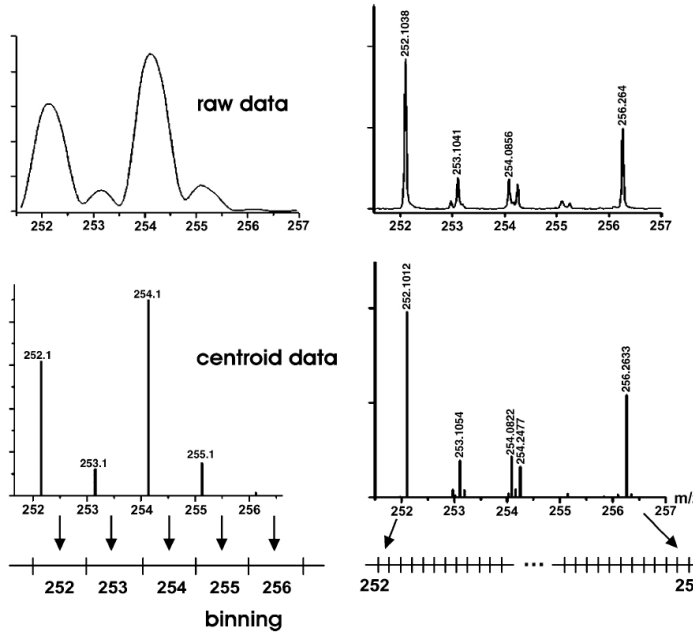

Figure 2.6 Mass spectra can be presented in two formats continuum (raw data) and centroid (compressed data). The continuum spectrum contains the maximum information and corresponds to the raw data, but takes up considerable storage space. The centroid spectrum is a compact and a processed variant of the continuum spectrum, which captures the major spectral information. Binning of centroid data is used to align spectra. Here examples of nominel (left) and high-resolution (right) data are shown. Taken from Smedsgaard and Nielsen (2005).

in the same bin, the binning will result in loss of information. For nominal mass spectra a bin width of 1 Da is appropriate, otherwise the bins can be designed using *a priori* knowledge about mass precision and data type (Smedsgaard and Nielsen, 2005).

When comparing multiple mass spectra the data need to be brought to the same scale by normalization in order to properly compare the data. There are varies ways of doing this. Sum normalization is done by dividing the intensity of the individual ions in the mass spectrum by the total sum of intensity. Alternatively, the maximum peak (base peak[4]) can be used to scale the spectrum. In case an internal standard or a marker is present in the spectra, its corresponding intensity can also be used for normalization.

From a series of mass spectra, *e.g.* obtained from a chromatographical run, the data are three-dimensional (time, mass and intensity), and different (two-dimensional) chromatograms can be extracted from the data:

Total ion chromatogram (TIC) is the sum of all ion intensities in each mass spectrum as a function of time

Base peak chromatogram (BPC) is the intensity of the base peak in each mass spectrum as a function of time

Extracted ion chromatogram (EIC) is the intensity of a specified ion as a function of time

References

N. B. Cech and C. G. Enke. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.*, 20(6):362–387, 2001.

R. B. Cole. Some tenets pertaining to electrospray ionization mass spectrometry. *J. Mass Spectrom.*, 35(7):763–772, 2000.

O. Fiehn, J. Kopka, R. N. Trethewey, and L. Willmitzer. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.*, 72(15):3573–3580, 2000.

M. W. Forbes, M. Sharifi, T. Croley, Z. Lausevic, and R. E. March. Simulation of ion trajectories in a quadrupole ion trap: a comparison of three simulation programs. *J. Mass Spectrom.*, 34(12):1219–1239, 1999.

A. Gomez and K. Q. Tang. Charge and fission of droplets in electrostatic sprays. *Physics of Fluids*, 6(1):404–414, 1994.

[4]In a mass spectrum the ion presenting the highest intensity is referred to as the base peak.

J. M. Halket, D. Waterman, A. M. Przyborowska, R. K. Patel, P. D. Fraser, and P. M. Bramley. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.*, 56(410):219–243, 2005.

K. R. Jonscher and J. R. Yates. The quadrupole ion trap mass spectrometer - A small solution to a big challenge. *Anal. Biochem.*, 244(1):1–15, 1997.

P. Kebarle. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom.*, 35(7):804–817, 2000.

P. Kebarle and L. Tang. From ions in solution to ions in the gas-phase - the mechanism of electrospray mass-spectrometry. *Anal. Chem.*, 65(22):A972–A986, 1993.

T. Kind and O. Fiehn. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*, 7:234, 2006.

R. E. March. An introduction to quadrupole ion trap mass spectrometry. *J. Mass Spectrom.*, 32(4):351–369, 1997.

B. K. Matuszewski, M. L. Constanzer, and C. M. Chavez-Eng. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Anal. Chem.*, 75(13):3019–3030, 2003.

D. V. McCalley. Comparison of peak shapes obtained with volatile (mass spectrometry-compatible) buffers and conventional buffers in reversed-phase high-performance liquid chromatography of bases on particulate and monolithic columns. *J. Chromatogr. A*, 987(1-2):17–28, 2003.

F. W. McLafferty and F. Turecek. *Interpretation of mass spectra*. University Science Books, Sausalito, 4th edition, 1993.

J. Smedsgaard and J. Nielsen. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J. Exp. Bot.*, 56(410):273–286, 2005.

C. Steel and M. Henchman. Understanding the quadrupole mass filter through computer simulation. *J. Chem. Educ.*, 75(8):1049–1054, 1998.

P. J. Taylor. Matrix effects: The Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clin. Biochem.*, 38(4):328–334, 2005.

Z. R. Wu, W. Q. Gao, M. A. Phelps, D. Wu, D. D. Miller, and J. T. Dalton. Favorable effects of weak acids on negative-ion electrospray ionization mass spectrometry. *Anal. Chem.*, 76(3):839–847, 2004.

Chapter 3

Metabolite profiling platforms in metabolomics

The degree of chemical heterogeneity increases from DNA to metabolites. DNA and RNA are polymers of 5 nucleotides and proteins are composed of 22 amino acids. Indeed proteins are much more diverse than DNA and RNA, but anyhow much more homogenous than metabolites. Particularly, when they are digested to peptides. Metabolites vary in chemical and physical properties from ionic and hydrophilic to lipophilic. This makes it challenging to cover the complete set of metabolites in one go.

This chapter will give an overview of metabolite analysis techniques in metabolomics. The emphasis will be on applications based on MS and a short introduction of nuclear magnetic resonance spectroscopy (NMR) is included, since NMR along with mass spectrometry (MS) are the two most prominent candidates to undertake the analytical challenges in metabolomics.

3.1 Fingerprinting

The principle of fingerprinting is to capture a snapshot of a metabolic extract and the fingerprinting technique is used to obtain a metabolic fingerprint (or footprint; see Table 1.1 for definitions). It is comparable with a picture developed from physical properties of the metabolites, *e.g.* mass, nuclear spin, bond vibration, light absorption *etc.* The data acquisition is fast and data can be reduced to a 2-dimensional space representing the physical property orthogonal to the magnitude. Some qualitative and quantitative information might be present in the data, but this is not a prerequisite for fingerprinting.

3.1.1 Direct infusion MS

Direct infusion MS is a mass spectral fingerprint of a sample injected directly into the ion source of a mass spectrometer. Some metabolites from the sample are ionized and enter the mass analyzer, in which they are separated according to mass (Chapter 2). The outcome of the analysis is a mass spectrum comprising the metabolite sample composition to a certain extend. The analysis time is short typically lies in the range of three to five minutes.

Direct infusion MS has successfully been used in different classification studies. Smedsgaard and colleagues (Smedsgaard, 1997; Smedsgaard et al., 2004; Smedsgaard and Nielsen, 2005) have used direct infusion MS to profile secondary metabolites from filamentous fungi for chemotaxonomy, where closely related species of *Penicillium* were classified. From the mass spectral data several secondary metabolites could be identified based on masses and isotope patterns. Similarly, direct infusion MS was used for identification of bacteria by analysis of crude cell extracts (Vaidyanathan et al., 2002). Here, five bacterial strains were studied and the predominant biomakers from the analysis were found to be phospholipids, glycolipids and proteins. In a functional genomic context, Allen et al. (2003) analyzed the extracellular metabolite profiles (footprints) of single knock-out strains of the yeast *Saccharomyces cerevisiae* by direct infusion MS. The results proved that mutants with related genotypes express similar extracellular metabolite profiles and the concept was applicable for assignment of gene functions by guilt-by-association. Extension of this strategy demonstrated that metabolic footprinting also can be used to determine mode-of-action of antifungal compounds (Allen et al., 2004).

Matrix effects may cause interference during direct infusion MS analysis as discussed in Section 2.1.2, and there is a risk of favoring metabolites with high ionization efficiency, since all metabolites are ionized together. However, samples obtained from similar chemical conditions can provide information about changes in metabolic states as discussed in Chapter 8. Quantitative analysis by direct infusion MS is rare, but Nagy et al. (2003) analyzed amino acids in blood spots by direct infusion tandem MS, where addition of isotope-labeled internal standards allowed quantification of 19 native amino acids.

Application of high resolution mass spectrometers can increase the chemical information from direct infusion MS, since they make it possible to determine the accurate mass of compounds. Even compositional different compounds that have the same unit mass can be separated on the mass scale, *e.g.* lysine ($C_6H_{14}N_2O_2$; M = 146.1055 Da) and glutamine ($C_5H_{10}N_2O_3$; M = 146.0691 Da), and from the accurate mass measurement the elemental composition can be directly computed (Smedsgaard and Nielsen, 2005; Aharoni et al., 2002; Sleno et al., 2005).

Conformational isomers can not be resolved by MS alone simply because they will have the same mass, but tandem MS or coupling of separation techniques in front of the MS can help this out. Combination of separation techniques with MS tremendously improves the resolution of complex samples, see Section 3.2.

3.1.2 Nuclear magnetic resonance spectroscopy

NMR is a versatile technology and probably the technique of choice when it comes to detailed structure elucidation of purified metabolites, but NMR is also widely applied for metabolic fingerprinting. NMR relies on the nuclear spin of atomic nuclei. The nuclear spin is posed by unpaired protons and neutrons and makes the nuclei act like small magnets and thus *e.g.* ^{1}H, ^{13}C, ^{15}N and ^{31}P can be analyzed by NMR.

The sensitivity is not as high as for MS; however, NMR does not discriminate between metabolites in the samples and thereby present an unbiased technique for metabolite fingerprinting that is quantitative even in complex mixtures. Nicholson and co-workers have pioneered the application of NMR for metabolite fingerprinting (Bollard et al., 2005; Lindon et al., 2003). In particular, they analyzed urine and blood by NMR and successfully identified biomarkers for diagnostics. Fingerprinting by ^{1}H-NMR of a metabolite extracts from yeast was proposed for assignment of unknown gene functions. The analysis of six yeast knock-out strains proved to classify and relate the genotypes by multivariate statistics, which potentially can be applied for functional genomics (Raamsdonk et al., 2001).

NMR is non-destructive and therefore *in vivo* analysis is also possible. Insertion of a special probe into the NMR instrument makes it possible to grow microbial cell in suspension and hereby perform non-invasive studies of the metabolite concentrations in a dynamic state (Gmati et al., 2005; Gonzalez et al., 2000; Neves et al., 2005; Weuster-Botz and de Graaf, 1996).

3.2 Metabolite profiling

Adding separation in time to the detection method brings us to metabolite profiling. This strategy provides increased resolution in comparison with fingerprinting, and the metabolites will be at least partly separated and detected individually. This adds another dimension to the data and the combination of a separation technology and MS gives a 2-dimensional separation in time and mass as illustrated in Figure 3.1 with intensity then being the third dimension.

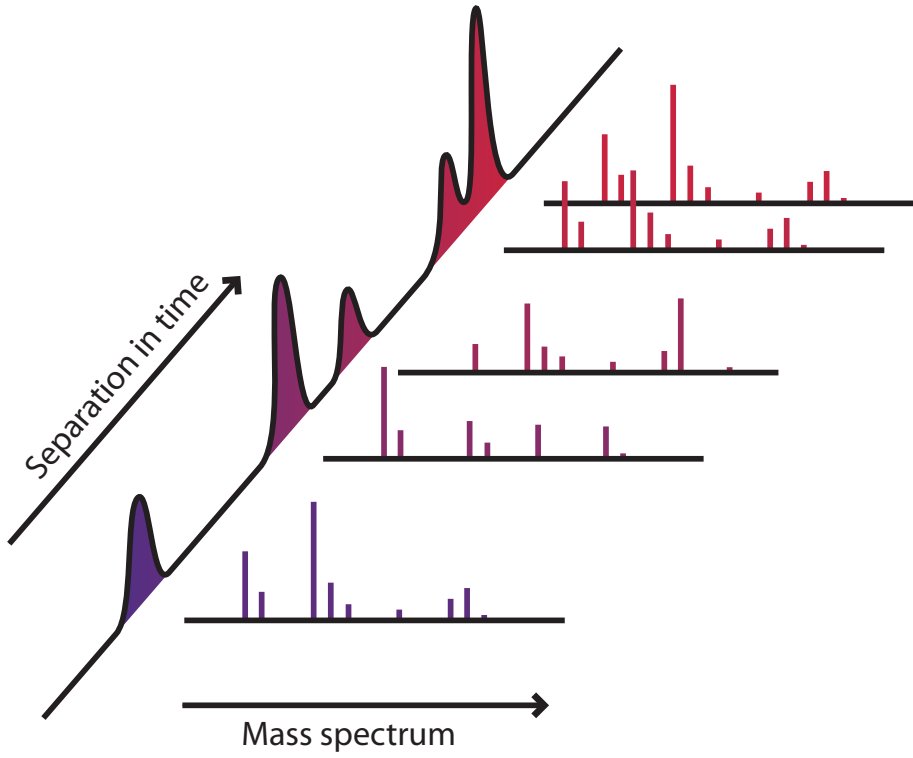

Figure 3.1 The combination of a separation technique and mass spectrometry provides a very powerful platform that enables high resolution and sensitive detection

3.2.1 Gas chromatography coupled to MS

Any chromatographic technique is based on distribution between a stationary phase and a mobile phase. In gas chromatography (GC), the mobile phase is a gas and therefore the molecules that are to be analyzed are required to be volatile enough to be evaporated to the mobil carrier, which normally is helium. GC can be used to analyze the volatile part of the metabolome and although this part might often be neglected, it does provide insight into metabolism and taxonomy (Larsen and Frisvad, 1994; Karlshoj and Larsen, 2005). On the other hand, many metabolites are small polar molecules that are not readily evaporated into the gas phase. This can be overcome by derivatization that converts the functional polar groups in the molecule into non-polar groups.

A classical derivatization reaction is silylation, where active hydrogen atoms

from hydroxyl, amine, thiol and carboxylic acid groups are substituted by silyl groups, *e.g.* trimethylsilyl [$-Si(CH_3)_3$] or tert-butyldimethylsilyl [$-Si(CH_3)_2C_4H_9$]. The reactions proceed under strictly anhydrous conditions at slightly elevated temperature and produces derivates that are generally less polar, more volatile, and thermally more stable than the parent compound. The derivatization reaction is very versatile and covers most primary metabolite although not all of them will end up being volatile (Halket and Zaikin, 2003; Koek et al., 2006; Roessner et al., 2000). Alternatively, alkylation can be used for derivatization of functional group with acidic or basic hydrogen atoms, *e.g.* carboxylic acids, and primary and secondary amines (Halket and Zaikin, 2004). Chloroformates is used for alkylation and have been suggested as a good alternative for silylation. They have successfully been applied in analysis of amino and organic acids in blood (Husek, 1995, 1998) and metabolite extracts (Villas-Boas et al., 2003, 2005). One of the major advantages of the chloroformate reaction is that it can be performed under partly aqueous conditions and at room temperature, but it does not enable analysis of sugars and sugar derived compounds as these are not derivatized. One thing to keep in mind is that the derivatization reaction rarely results in a single product, but multiple byproducts may be formed. This complicates the chromatograms and makes the data analysis more cumbersome, since one peak does not necessarily correspond to one metabolite.

Coupling of capillary GC to MS combines two strong technologies. GC offers high separation efficiency between even similar compounds whereas MS produces compound-specific mass spectra that may resolve co-eluting compounds. Metabolite profiling of plants has significantly contributed to the field of metabolite profiling by GC-MS (Fiehn et al., 2000; Roessner et al., 2000, 2001) and the methodologies can readily be transferred to microbial metabolite profiling (Strelkov et al., 2004). Recently, two-dimensional GC (GC×GC)-TOF-MS instruments have been applied for metabolite profiling and this is currently the most comprehensive GC-based technology available. Two-dimensional GC enables very high chromatographic resolution and high scan rates from the TOF mass spectrometer makes it possible to resolve very complex samples with high sensitivity (Kell et al., 2005; Jover et al., 2005; Adahchour et al., 2006). The data are further resolved compared to conventional GC-MS and GC×GC-MS can be considered as 4D-data.

3.2.2 Liquid chromatography coupled to MS

Introduction of atmospheric pressure ionization (API) techniques opened up new possibilities for metabolite profiling. The coupling of liquid chromatography (LC) to MS became reality and LC-MS instruments are nowadays standard equipment

in many labs. Till today the major applications of LC-MS have been within environmental analysis, analysis of pharmaceuticals, and profiling of biofluids. Although LC-MS has yet not been as widely used for primary metabolite profiling, it has a major potential to complement GC-MS. Derivatization is not required and low separation temperatures compared to GC-MS reduces the risk of degradation of heat labile compounds.

Separation of highly polar and ionic metabolites can be achieved by ion-exchange chromatography that includes gradient elution at high salt concentrations. This is not readily compatible with ESI, but insertion of a desalting device in between the LC and the MS is a solution for improved sensitivity and robustness. Reversed phase ion-pair chromatography is another option for separation of ionic metabolites. The only requirement is that the counter-ion used for ion-pairing should be volatile *e.g.* alkyl amines and perfluoro-carboxylic acids, even though these might still present a problem with contamination of the ion-source.

Matrix effects in the ESI process (Section 2.1.2) are especially critical for quantitative studies. Although tandem MS allows highly specific and sensitive detection of metabolites, matrix effects during ionization are obstructive for quantification. Addition of isotopic-labeled internal standards can overcome the discrimination observed during ionization and extend the dynamic range for quantification. The labeled standard is ionized together with the compound of interest and the ions will be separated by the mass spectrometer (Stokvis et al., 2005). The only drawback is that only a limited number of labeled metabolites are commercially available and they tend to be expensive; but *in vivo* synthesis by feeding labeled substrates is an option (Birkemeyer et al., 2005; Mashego et al., 2004; Wu et al., 2005).

Metabolites from the central carbon metabolism can be directly analyzed by LC-MS (Buchholz et al., 2001; van Dam et al., 2002; Huck et al., 2003) and also nucleotides (Coulier et al., 2006; Xing et al., 2004). Lu et al. (2006) has developed a LC-MS method for quantitative analysis of nitrogen-containing compound covering 90 metabolites and using this method it was possible to quantify 36 metabolites from *Salmonella enterica*. The continuing development of new stationary phases for LC columns are promising and will pose new possibilities through novel stationary phases with improved resolution. One of the most recent technologies is ultrahigh pressure liquid chromatography. Here reduction of the particle size of the column material to sub-2 μm and high linear flow rates have significantly improved the chromatographic resolution to levels previously only seen for GC (Plumb et al., 2004).

3.2.3 Capillary electrophoresis coupled to MS

A third separation technique is capillary electrophoresis (CE). Separation by CE relies on differential migration of ions in an electrical field and is performed in a buffer-filled capillary of fused silica. The capillary is placed in two separate buffer reservoirs with a potential difference of up to 30 kV. The potential over the capillary mediates an electroosmotic flow that carries the analytes along while they are separated by differential migration. The electroosmotic flow can be varied by changing the applied potential and the ion strength in the buffer, but the flow is typically in the range of nL/min. The migration velocity is determined by the charge-to-volume ratio of the ion and the overall migration velocity will be a sum of the electroosmostic flow and the migration of the individual ions.

CE offers high separation efficiencies, which makes it powerful for analysis of complex samples (Ramautar et al., 2006) although sensitivity might be limited by fairly small injection volumes in the range of nL. Coupling of CE to mass spectrometry is possible, but currently not that widespread. Due to the low flow rates the ionization is mostly performed by ESI, and the interface between the CE instrument and the ESI source often includes a coaxial sheath flow to maintain a stable spray (Schmitt-Kopplin and Frommberger, 2003).

CE is ideal for separation of ionic metabolites. There is no requirement for derivatization and many different compound classes can be analyzed *e.g.* organic acids, amino acids, nucleotides and carbohydrates with indirect UV detection (Soga and Imaizumi, 2001). Soga et al. (2002) coupled the CE to MS which enabled analysis of 32 central anionic metabolites from glycolysis and TCA cycle. Later, a combination of three different CE-MS methods was used to cover the cationic metabolites, the anionic metabolites, and nucleotides and coenzymes, respectively, which enabled detection of more than 1,500 compounds from a *Bacillus subtilis* extract. Of these compounds, 150 were identified based analysis of 352 metabolite standards, and 83 unknown compounds were assigned (Soga et al., 2003).

References

M. Adahchour, J. Beens, R. J. J. Vreuls, and U. A. T. Brinkman. Recent developments in comprehensive two-dimensional gas chromatography (GC×GC) I. Introduction and instrumental set-up. *Trends Anal. Chem.*, 25(5):438–454, 2006.

A. Aharoni, C. H. R. de Vos, H. A. Verhoeven, C. A. Maliepaard, G. Kruppa, R. Bino, and D. B. Goodenowe. Nontargeted metabolome analysis by use of fourier transform ion cyclotron mass spectrometry. *OMICS*, 6(3):217–234, 2002.

J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.*, 21(6):692–696, 2003.

J. Allen, H. M. Davey, D. Broadhurst, J. J. Rowland, S. G. Oliver, and D. B. Kell. Discrimination of modes of action of antifungal substances by use of metabolic footprinting. *Appl. Environ. Microbiol.*, 70(10):6157–6165, 2004.

C. Birkemeyer, A. Luedemann, C. Wagner, A. Erban, and J. Kopka. Metabolome analysis: the potential of *in vivo* labeling with stable isotopes for metabolite profiling. *Trends Biotechnol.*, 23(1):28–33, 2005.

M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.*, 18(3):143–162, 2005.

A. Buchholz, R. Takors, and C. Wandrey. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. *Anal. Biochem.*, 295(2):129–137, 2001.

L. Coulier, R. Bas, S. Jespersen, E. Verheij, M. J. van der Werf, and T. Hankemeier. Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.*, 78(18):6573–6582, 2006.

O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R. N. Trethewey, and L. Willmitzer. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.*, 18(11):1157–1161, 2000.

D. Gmati, J. K. Chen, and M. Jolicoeur. Development of a small-scale bioreactor: Application to *in vivo* NMR measurement. *Biotechnol. Bioeng.*, 89(2):138–147, 2005.

B. Gonzalez, A. de Graaf, M. Renaud, and H. Sahm. Dynamic *in vivo* P-31 nuclear magnetic resonance study of *Saccharomyces cerevisiae* in glucose-limited chemostat culture during the aerobic-anaerobic shift. *Yeast*, 16(6):483–497, 2000.

J. M. Halket and V. G. Zaikin. Derivatization in mass spectrometry – 1. Silylation. *Eur. J. Mass Spectrom.*, 9(1):1–21, 2003.

J. M. Halket and V. G. Zaikin. Derivatization in mass spectrometry – 3. Alkylation (arylation). *Eur. J. Mass Spectrom.*, 10(1):1–19, 2004.

J. H. Huck, E. A. Struys, N. M. Verhoeven, C. Jakobs, and M. S. van der Knaap. Profiling of pentose phosphate pathway intermediates in blood spots by tandem mass spectrometry: application to transaldolase deficiency. *Clin. Chem.*, 49(8):1375–1380, 2003.

P. Husek. Simultaneous profile analysis of plasma amino and organic acids by capillary gas chromatography. *J. Chromatogr. B*, 669(2):352–357, 1995.

P. Husek. Chloroformates in gas chromatography as general purpose derivatizing agents. *J. Chromatogr. B*, 717(1-2):57–91, 1998.

E. Jover, M. Adahchour, J. M. Bayona, R. J. J. Vreuls, and U. A. T. Brinkman. Characterization of lipids in complex samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. *J. Chromatogr. A*, 1086(1-2): 2–11, 2005.

K. Karlshoj and T. O. Larsen. Differentiation of species from the *Penicillium roqueforti* group by volatile metabolite profiling. *Journal of Agricultural and Food Chemistry*, 53(3):708–715, 2005.

D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic, and S. G. Oliver. Metabolic footprinting and systems biology: The medium is the message. *Nat. Rev. Microbiol.*, 3(7):557–565, 2005.

M. M. Koek, B. Muilwijk, M. J. van der Werf, and T. Hankemeier. Microbial metabolomics with gas chromatography/mass spectrometry. *Anal. Chem.*, 78(4): 1272–1281, 2006.

T. O. Larsen and J. C. Frisvad. A simple method for collection of volatile metabolites from fungi based on diffusive sampling from petri dishes. *J. Microbiol. Meth.*, 19(4): 297–305, 1994.

J. C. Lindon, E. Holmes, and J. K. Nicholson. So whats the deal with metabonomics? Metabonomics measures the fingerprint of biochemical perturbations caused by disease, drugs, and toxins. *Anal. Chem.*, 75(17):384A–391A, 2003.

W. Lu, E. Kimball, and J. D. Rabinowitz. A high-performance liquid chromatography-tandem mass spectrometry method for quantitation of nitrogen-containing intracellular metabolites. *J. Am. Soc. Mass Spectrom.*, 17:37–50, 2006.

M. R. Mashego, L. Wu, J. C. van Dam, C. Ras, J. L. Vinke, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen. MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnol. Bioeng.*, 85(6):620–628, 2004.

K. Nagy, Z. Takats, F. Pollreisz, T. Szabo, and K. Vekey. Direct tandem mass spectrometric analysis of amino acids in dried blood spots without chemical derivatization for neonatal screening. *Rapid Commun. Mass Spectrom.*, 17(9):983–990, 2003.

A. R. Neves, W. A. Pool, J. Kok, O. P. Kuipers, and H. Santos. Overview on sugar metabolism and its control in *Lactococcus lactis* - The input from *in vivo* NMR. *FEMS Microbiol. Rev.*, 29(3):531–554, 2005.

R. Plumb, J. Castro-Perez, J. Granger, I. Beattie, K. Joncour, and A. Wright. Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, 18(19):2331–2337, 2004.

L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.*, 19(1):45–50, 2001.

R. Ramautar, A. Demirci, and G. J. de Jong. Capillary electrophoresis in metabolomics. *Trends Anal. Chem.*, 25(5):455–466, 2006.

U. Roessner, C. Wagner, J. Kopka, R. N. Trethewey, and L. Willmitzer. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.*, 23(1):131–142, 2000.

U. Roessner, A. Luedemann, D. Brust, O. Fiehn, T. Linke, L. Willmitzer, and A. Fernie. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, 13(1):11–29, 2001.

P. Schmitt-Kopplin and M. Frommberger. Capillary electrophoresis - mass spectrometry: 15 years of developments and applications. *Electrophoresis*, 24(22-23):3837–3867, 2003.

L. Sleno, D. A. Volmer, and A. G. Marshall. Assigning product ions from complex MS/MS spectra: The importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.*, 16(2):183–198, 2005.

J. Smedsgaard. Micro-scale extraction procedure for standardized screening of fungal metabolite production in cultures. *J. Chromatogr. A*, 760(2):264–270, 1997.

J. Smedsgaard and J. Nielsen. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J. Exp. Bot.*, 56(410):273–286, 2005.

J. Smedsgaard, M. E. Hansen, and J. C. Frisvad. Classification of terverticillate Penicillia by electrospray mass spectrometric profiling. *Stud. Mycol.*, 49:243–251, 2004.

T. Soga and M. Imaizumi. Capillary electrophoresis method for the analysis of inorganic anions, organic acids, amino acids, nucleotides, carbohydrates and other anionic compounds. *Electrophoresis*, 22(16):3418–3425, 2001.

T. Soga, Y. Ueno, H. Naraoka, Y. Ohashi, M. Tomita, and T. Nishiokat. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal. Chem.*, 74 (10):2233–2239, 2002.

T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, and T. Nishioka. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.*, 2(5):488–494, 2003.

E. Stokvis, H. Rosing, and J. H. Beijnen. Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not? *Rapid Commun. Mass Spectrom.*, 19(3):401–407, 2005.

S. Strelkov, M. von Elstermann, and D. Schomburg. Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. *Biol. Chem.*, 385(9):853–861, 2004.

S. Vaidyanathan, D. B. Kell, and R. Goodacre. Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. *J. Am. Soc. Mass Spectrom.*, 13(2):118–128, 2002.

J. C. van Dam, M. R. Eman, J. Frank, H. C. Lange, G. W. van Dedem, and S. J. Heijnen. Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. *Anal. Chim. Acta*, 460:209–218, 2002.

S. G. Villas-Boas, D. G. Delicado, M. Akesson, and J. Nielsen. Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry. *Anal. Biochem.*, 322(1):134–138, 2003.

S. G. Villas-Boas, J. F. Moxley, M. Akesson, G. Stephanopoulos, and J. Nielsen. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem. J.*, 388:669–677, 2005.

D. Weuster-Botz and A. A. de Graaf. Reaction engineering methods to study intracellular metabolite concentrations. In *Advances in Biochemical Engineering/Biotechnology*, volume 54, pages 75–108. Springer Berlin/Heidelberg, 1996.

L. Wu, M. R. Mashego, J. C. van Dam, A. M. Proell, J. L. Vinke, C. Ras, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly C-13-labeled cell extracts as internal standards. *Anal. Biochem.*, 336(2):164–171, 2005.

J. Xing, A. Apedo, A. Tymiak, and N. Zhao. Liquid chromatographic analysis of nucleosides and their mono-, di- and triphosphates using porous graphitic carbon stationary phase coupled with electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.*, 18(14):1599–1606, 2004.

Chapter 4

Data analysis tools

From metabolite profiling, ideally all metabolites should be quantified. This gives – for yeast at least – 600 or so metabolites corresponding to 600 measurements/variables per sample point placing the sample in a hyper-dimensional space. Among the multiple variables probably only a few are needed to describe the investigated case adequately. Thus, the major challenge is how to pick out and identify these few variables of importance.

The pipeline for data analysis can be divided in three stages (Brown et al., 2005):

1. Preprocessing
2. Data reduction
3. Data analysis

First, the raw data need to be preprocessed and converted into a suitable form, *e.g.* improvement of signal-to-noise (S/N), normalization, deconvolution and feature extraction. For most multivariate data mining techniques, the data need to be aligned and conveniently organized into a matrix or an array. In a matrix-format the n rows will contain the samples and the m columns the variables resulting in a $n \times m$-matrix. Likewise, the metadata associated with the samples can be organized into a similar matrix containing n rows of samples and a number of columns representing the metadata. The metadata could correspond to experimental conditions like carbon source, group information, genotype, strain, sampling time *etc.*

The data reduction stage aims at qualifying the data for further analysis. Here, outliers are detected and excluded, and replicate samples are used to confirm that

features/variables are conserved within the replicates. A selection of variables for description of group information can be done through statistical testing to identify the presence of group characteristics. Student's t-test and one-way analysis of variance (ANOVA) are common candidates for selection of variables that describe group differences. The tests return a p-value[1], which will indicate the ability of the variable to describe group differences.

Finally, the data analysis should reveal data patterns that represent biological information and can be used to generate hypotheses for further studies.

4.1 Metrics for mass spectral comparison

A binned mass spectrum (see Section 2.4) can be represented as a vector containing the spectral ion intensities ordered according to the mass scale. Distance (or the inverse similarity) metrics can then be used for comparison of mass spectra. Stein and Scott (1994) compared five search algorithms for identification of compounds from mass spectral libraries. From the study, Stein and Scott (1994) concluded that the dot-product algorithm performed best for identification of compounds in GC-MS data. Later, Stein (1999) published the algorithm underlying AMDIS (automated mass spectral deconvolution and identification system) for analysis GC-MS data. Here the dot-product was also applied for measure of spectral similarity. Mathematically, the normalized dot-product is computed according to Equation 4.1 including weighting by the ion mass (m) and scaling of intensity/abundance by the power of $\frac{1}{2}$ (Stein, 1999). Equation 4.1 takes values from 0 to 100, where 100 corresponds to the similarity between identical spectra.

$$\text{similarity index} = 100\frac{\left(\sum_i m_i \left[A_{u,i}A_{r,i}\right]^{\frac{1}{2}}\right)^2}{\sum_i A_{u,i}m_i \ \sum_i A_{r,i}m_i} \tag{4.1}$$

where $A_{u,i}$ and $A_{r,i}$ are the abundances of the ith ion in the user and the reference spectrum, respectively, and m_i is the mass of the ith ion.

The method described above rely on aligned and binned data. For accurate mass data, binning of mass spectra for alignment results in loss of information, since the accurate mass is no longer preserved in the data structure and adjacent ions with mass differences within bin width might be confounded as they can fall in the same bin. Hansen and Smedsgaard (2004) solved this problem by defining a distance for accurate mass spectra on a continuous scale. The distance measure is computed between two spectra by first pairing the ions with the same mass

[1] The p-value is the probability of obtaining a samples that is at least as extreme/"impressive" as that obtained, assuming that the null hypothesis is true, so that the finding was the result of chance alone.

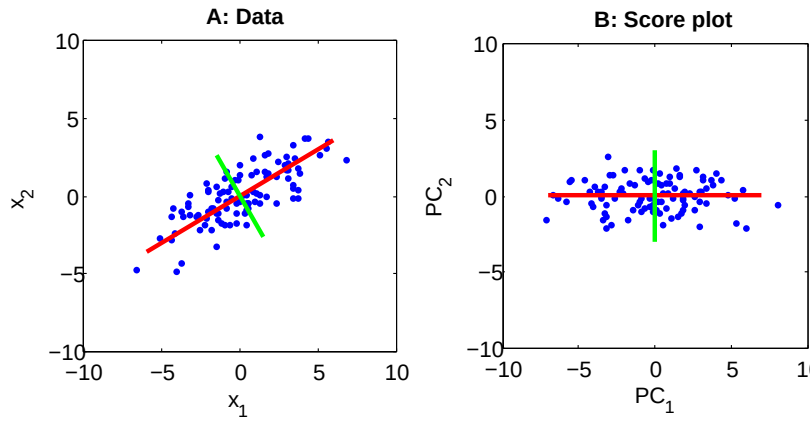

Figure 4.1 Principal component analysis (PCA) is a spacial rotation of the data space in order to present maximum data variation derived from covariation. For illustration, a 2-dimensional data set (**A**) is rotated by PCA, which results in the score-plot (rotated space) in **B**. The red and green line represent the first and second principal component (PC), respectively. It is illustrated that the major variation is captured by the first PC.

between the spectra and then calculating the pairwise distance of these by the Jeffreys-Matusitas distance. Combining these pairwise distances and weighting them according to relative number of paired ions give the accurate mass spectrum distance between two spectra. The performance of the algorithm was tested against existing library search algorithms, and it was found to perform at least as good as these (Hansen and Smedsgaard, 2004).

4.2 Reduction of dimensionality

Reduction of dimensionality is applied in order to get an overview and visualization of the data. It is a common scenario in biology that the number of samples is relatively low compared to the number of variables available and sometimes the variables are also highly correlated[2].

Principal component analysis (PCA) is a popular method for visualization of metabolome data through reduction of dimensionality. The method is unsupervised which means that no sample information (metadata) is included in the analysis. In principle, the PCA is a spacial rotation by linear transformation of

[2]Specifically for mass spectrometry, the presence of isotopes gives rise to redundant information in mass data, because the intensity of the isotope ions is strictly correlated due to the natural isotope abundances.

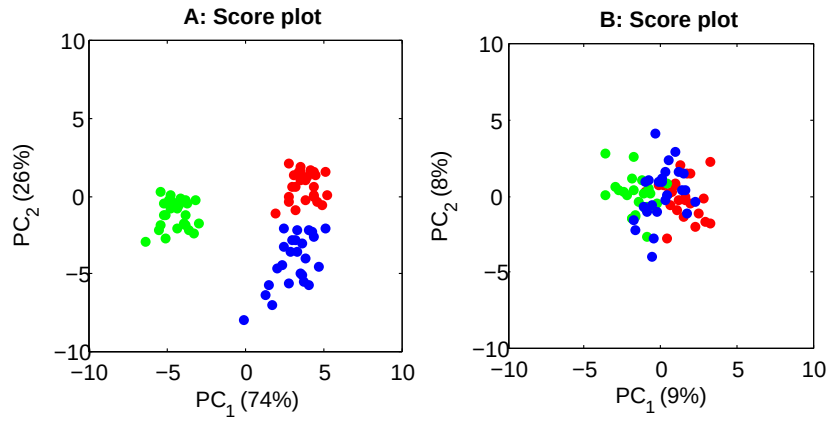

Figure 4.2 Principal component analysis (PCA) is affected by noise, which can impact the ability of the analysis to separate groups even though group information is available in the data. Using two variables segregating three group, PCA efficiently separates the two groups by the first principal component (**A**). The performance of PCA is weakened when adding noise by including 25 variables with random values in addition to the two variables. Then the PCA is no longer capable of clearly separating the groups (**B**).

the variable space in such way that it represents the maximum variation in the data set. The computed principal components (PCs) are organized in such way that the first PC accounts for the maximum variation, the second PC describes the second most variation *etc.* (Figure 4.1). All PCs are orthogonal to each other and the maximum number of PCs being the smaller of $n-1$ and m. This way, the PCs describing the major variation, *e.g.* 95% or 99%, can be selected to represent the data in a lower dimensional space.

Before PCA, the data matrix is normally centered and scaled (Bro and Smilde, 2003) although Allen et al. (2003) stated that this can cloud the underlying statistical trends, but the variables in the data are not corrected for magnitude by omitting this step. The PCA can be done by eigenvalue decomposition of the data covariance matrix or by singular value decomposition of the centered and scaled data matrix. This way the unit eigenvectors (or singular vectors) correspond to the PCs, which are numbered according to descending eigenvalue. In relation to PCA the PCs are referred to as loadings and represent the weights of the original variables in the linear transformation. Multiplication of the loading matrix and the data matrix will give the scores, which are the coordinates in the new projection space that give the transformation illustrated in Figure 4.1 (Jolliffe, 2002).

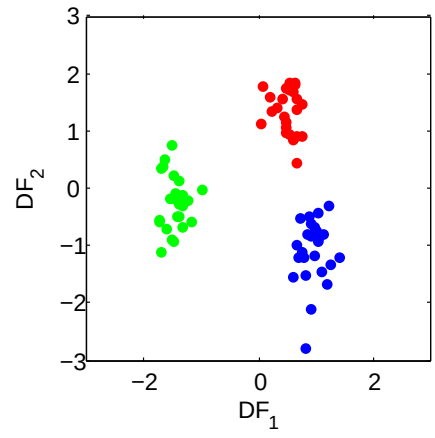

Figure 4.3 Fisher discriminant analysis (FDA) of the same noisy data as were used for the PCA score plot in Figure 4.2B.

The widespread use of PCA is probably mediated by the easy implementation of the algorithm that needs no parameters. For PCA to be successful in pattern visualization, the major variation should capture the group segregation; however, the objective of variance maximization does not always correspond to information. The PCA is affected by noisy variables, which blur the analysis when identification of group information is intended (Halouska and Powers, 2006). This is also illustrated in Figure 4.2, where two variables are describing differences between three groups. Applying the PCA on the data matrix containing the two variables yields a spacial rotation where the three groups are separated (Figure 4.2A). Adding another 25 variables with random values to the data matrix and performing a PCA gives a more blurred picture (Figure 4.2B). This is simply because the PCA maximizes variation and does not consider group information.

Similar to the PCA is Fisher (or linear) discriminant analysis (FDA) that also is used to project data into a lower dimensional space. Where PCA maximizes variation, the FDA instead maximizes between-group variation while minimizing the within-group variation. Hereby turning FDA into a supervised method that uses group information (metadata) to compute the between- and within-group variation. FDA is better at overcoming noise in data, since it uses group information for the projection. In Figure 4.3 FDA is applied on the same data set as was used to obtain the PCA score plot in Figure 4.2B. Here, it is evident that the FDA is capable of capturing the group information and neglecting the noise. The FDA score plot (Figure 4.3) corresponds well to the PCA score plot in Fig-

ure 4.2A with the two variable describing the three groups. Actually, the FDA seems to give slightly less disperse clusters, which most likely is because FDA has a tendency of over-fitting the data when the group size is small compared the number of variables. Therefore it is always recommended to use cross-validation for FDA models or any other supervised methods.

Stephanopoulos et al. (2002) demonstrated the use of FDA in combination with ANOVA for microarray data analysis, and Raamsdonk et al. (2001) and Allen et al. (2003) applied the FDA in combination with PCA for metabolic finger- and footprinting data. Raamsdonk et al. (2001) and Allen et al. (2003) first aimed at reducing the variable space to describe 95% of the data variation by PCA and then the reduced space was subjected to discriminant analysis. This resulted in clustering of similar mutants of *Saccharomyces cerevisiae*. The only downside of the PCA-FDA approach is that it is difficult to backtrack and interpret the influence and loadings of the original variables.

References

J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.*, 21(6):692–696, 2003.

R. Bro and A. K. Smilde. Centering and scaling in component analysis. *J. Chemometr.*, 17(1):16–33, 2003.

M. Brown, W. B. Dunn, D. I. Ellis, R. Goodacre, J. Handl, J. D. Knowles, S. O'Hagan, I. Spasic, and D. B. Kell. A metabolome pipeline: from concept to data to knowledge. *Metabolomics*, 1(1):39–51, 2005.

S. Halouska and R. Powers. Negative impact of noise on the principal component analysis of NMR data. *J. Magnetic Resonance*, 178:88–95, 2006.

M. E. Hansen and J. Smedsgaard. A new matching algorithm for high resolution mass spectra. *J. Am. Soc. Mass Spectrom.*, 15(8):1173–1180, 2004.

I. T. Jolliffe. *Principal component analysis*. Springer Verlag, New York, 2nd edition, 2002.

L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.*, 19(1):45–50, 2001.

S. E. Stein. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.*, 10 (8):770–781, 1999.

S. E. Stein and D. R. Scott. Optimization and testing of mass-spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.*, 5(9):859–866, 1994.

G. Stephanopoulos, D. Hwang, W. A. Schmitt, J. Misra, and G. Stephanopoulos. Mapping physiological states from microarray expression measurements. *Bioinformatics*, 18(8):1054–1063, 2002.

Chapter 5

Global metabolite analysis of yeast: evaluation of sample preparation methods

Silas G. Villas-Bôas, Jesper Højer-Pedersen, Mats Åkesson, Jørn Smedsgaard, and Jens Nielsen

Yeast 22: 1155–1169, 2005.

Yeast
Yeast 2005; **22**: 1155–1169.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/yea.1308

Yeast Functional Analysis Report

Global metabolite analysis of yeast: evaluation of sample preparation methods

Silas G. Villas-Bôas[§#], Jesper Højer-Pedersen[§], Mats Åkesson[†], Jørn Smedsgaard and Jens Nielsen*
Centre for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Building 223, DK-2800 Kgs. Lyngby, Denmark

**Correspondence to:*
Jens Nielsen, Centre for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Building 223, DK-2800 Kgs, Lyngby, Denmark.
E-mail: jn@biocentrum.dtu.dk

[§]*Both authors contributed equally to this work.*

[#]*Present address: AgResearch Ltd, Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand.*

[†]*Present address: BioProcess Laboratories, Novo Nordisk A/S, Novo Allé, DK-2880 Bagsvaerd, Denmark.*

Received: 20 October 2004
Accepted: 30 August 2005

Abstract

Sample preparation is considered one of the limiting steps in microbial metabolome analysis. Eukaryotes and prokaryotes behave very differently during the several steps of classical sample preparation methods for analysis of metabolites. Even within the eukaryote kingdom there is a vast diversity of cell structures that make it imprudent to blindly adopt protocols that were designed for a specific group of microorganisms. We have therefore reviewed and evaluated the whole sample preparation procedures for analysis of yeast metabolites. Our focus has been on the current needs in metabolome analysis, which is the analysis of a large number of metabolites with very diverse chemical and physical properties. This work reports the leakage of intracellular metabolites observed during quenching yeast cells with cold methanol solution, the efficacy of six different methods for the extraction of intracellular metabolites, and the losses noticed during sample concentration by lyophilization and solvent evaporation. A more reliable procedure is suggested for quenching yeast cells with cold methanol solution, followed by extraction of intracellular metabolites by pure methanol. The method can be combined with reduced pressure solvent evaporation and therefore represents an attractive sample preparation procedure for high-throughput metabolome analysis of yeasts. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords: extraction; metabolome; metabolomics; quenching; sample preparation; freeze-drying

Introduction

Metabolomics and metabolome analysis is a broadly used term for analysis of a large number of metabolites from a single organism. In metabolome analysis of microbial cells the separation of intra- and extracellular metabolites is often required to gain insight into the control of metabolic pathways. Therefore, it is important to determine the intracellular levels of metabolites accurately, which requires efficient and reliable methods for sample preparation. There have been considerable efforts to develop sensitive and accurate methods for detection of metabolites (Buchholz *et al*., 2002; Soga *et al*., 2002; van Dam *et al*., 2002; Mashego *et al*., 2004; Villas-Bôas *et al*., 2005; and others) and to improve the efficiency of data mining and data analysis (Goodacre *et al*., 2004; Kell, 2004; van der Greef *et al*., 2004) but relative little attention has been given to sample preparation procedures.

Figure 1 summarizes the main steps in sample preparation for metabolome analysis. Since the metabolite levels reflect the ultimate response of a biological system to genetic or environmental changes, rapid stopping (quenching) of the cell metabolism and extracellular enzymatic activity is the first step for such methods. Metabolite concentrations are very prone to changes induced by (unnoticed) variation in the environment of the cell (de Koning and van Dam, 1992). The quenching procedure aims to stop cellular metabolism and prevent the turnover of metabolites, maintaining the *in vivo* metabolite concentrations at a constant

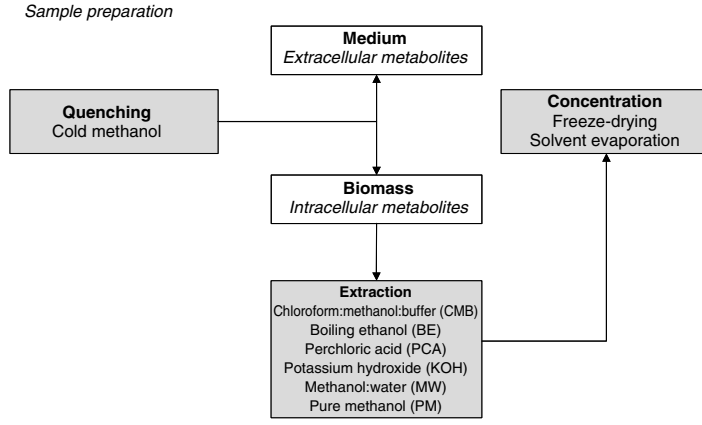

Figure 1. The main steps in sample preparation in connection with metabolome analysis of microorganisms. Grey squares show the steps evaluated in this work, listing the different methods examined

level. A rapid stop of metabolism is usually achieved through fast changes in temperature or pH. A limitation for quenching a microbial cell culture is the inherently associated dilution of the sample, which is due to the unfavourable ratio between biomass and the aqueous medium. The commonly used method is based on rapid sampling into a cold aqueous methanol solution, as proposed by de Koning and van Dam (1992). This method is preferred in microbial research because it is currently the only method that enables a reliable separation of intra- and extracellular metabolites while stopping cell metabolism. The cells are usually separated from the quenching mixture by centrifugation, requiring sample handling at low temperatures to eliminate or at least minimize turnover of metabolites prior to cell extraction.

A key question in connection with quenching yeast cell metabolism with cold methanol solution is whether it causes leakage of metabolites from the cells. Gonzalez *et al*. (1997) found significant amounts of glucose 6-phosphate, ATP and NAD^+ in the supernatant from methanol-quenched cells. However, the amount of these metabolites was roughly equal to those found in the culture medium. Hans *et al*. (2001) could not detect any amino acids in the supernatant of methanol-quenched yeast cells. In contrast to yeast cells, the intracellular metabolites from bacterial cells leak from the cells when these are brought into contact with cold methanol solution (60% v/v) and several intracellular metabolites have been found in the medium after the quenching process (Jensen *et al*., 1999; Letisse and Lindley, 2000; Wittmann *et al*., 2004).

Following quenching, the second step in the sample preparation is the separation of the extracellular metabolites from the biomass and subsequent extraction of the intracellular metabolites, achieving minimal losses by degradation and avoiding further biochemical reactions. The extraction step aims to induce complete leakage of the intracellular metabolites from the cells — and to some extent from its sub-compartments — into a liquid solution. Cell extraction makes the intracellular metabolites accessible to the different analytical methods employed. However, it is virtually impossible to avoid losses, mainly due to the diverse chemistry plus physical properties of the metabolome.

Again, de Koning and van Dam (1992) adapted a methodology originally designed for extraction of total lipids from animal tissues (Folch *et al*., 1957), based on buffered methanol–water mixture and chloroform, to extract polar metabolites from a yeast cell suspension. Their method has recently been applied for intracellular metabolite analysis of microorganisms (Cremin *et al*., 1995; Ruijter and Visser, 1996; Smits *et al*., 1998; Jensen *et al*., 1999; Zaldivar *et al*., 2002; Maharjan and Ferenci,

2003) and animal tissues (Le Belle *et al*., 2002). The method presents good recovery of phosphorylated compounds (Smits *et al*., 1998; Jensen *et al*., 1999). However, the recovery for other classes of metabolites has not been reported so far. In addition, this method is considered tedious and time-consuming, and uses chloroform, an undesirable solvent due to its toxic and carcinogenic effects.

Another popular extraction method for intracellular metabolites of yeast is based on the use of boiling ethanol as first described by Entian *et al*. (1977) and later revised by Gonzalez *et al*. (1997). It is widely used for the extraction of intracellular metabolites of microorganisms (Tweeddale *et al*., 1998; Hans *et al*., 2001; Lange *et al*., 2001; Visser *et al*., 2002; Castrillo *et al*., 2003). According to Gonzalez *et al*. (1997), the method is simple, fast, accurate and reliable. Glucose 6-phosphate, ATP, pyruvate, NAD^+ and NADH seem to be very stable in buffered boiling ethanol, presenting an excellent recovery after extraction. However, for metabolome analysis, this method is not free of problems, since not all metabolites are stable at the high temperature applied (Maharjan and Ferenci, 2003).

Recently, Maharjan and Ferenci (2003) proposed a method where extraction of intracellular metabolites from bacterial cells was achieved in cold methanol solution (50% v/v) after one freeze–thaw cycle in dry ice. It seems to present all the advantages of the boiling ethanol extraction method, in addition to preventing thermodegradation of the heat-labile compounds. However, this method has not been applied to metabolite extraction of organisms other than bacteria.

Classical acidic and alkalinic extraction methods are commonly used for the extraction of acid and alkaline stable compounds, respectively. These methods have been widely used in the extraction of metabolites from animal and plant tissues (Shryock *et al*., 1986; Kopka *et al*., 1995; Chen *et al*., 1998; Adams *et al*., 1999; Bouchereau *et al*., 2000; Conneely *et al*., 2002) and microorganisms (Cordeiro and Freire, 1996; Weuster-Botz, 1997; Hajjaj *et al*., 1998; Schaefer *et al*., 1999; Vaseghi *et al*., 1999; Letisse and Lindley, 2000; Martins *et al*., 2001; Buchholz *et al*., 2002; Buziol *et al*., 2002; Maharjan and Ferenci, 2003). Acidic extraction is commonly used for the extraction of nucleotides and water-soluble metabolites (Vaseghi *et al*., 1999; Martins *et al*., 2001; Buchholz *et al*., 2002), but for metabolome studies it is evident that not all metabolites are stable at the extreme pH, and destruction of some nucleotides has been reported during acidic extractions (Hajjaj *et al*., 1998).

Furthermore, most of the extraction methods result in a high degree of sample dilution, thus further diluting the already low concentration of some intracellular metabolites. Very often a third step in sample preparation is required, the concentration of the extract (Figure 1). Lyophilization, or freeze-drying, is commonly used to remove water from aqueous samples in order to concentrate and avoid thermal degradation. This method combines the advantage of both deep-freezing and dehydration. Alternatively, non-aqueous extracts can be concentrated by solvent evaporation. It should be noted, however, that both methods are unsuitable for volatile metabolites.

Therefore, this work aimed to evaluate the main steps in samples preparation for metabolome analysis of yeasts. We first examined the possible variations in the quenching step, using both buffered and non-buffered cold methanol solution in order to determine whether or not intracellular metabolite leakage takes place during this process. Second, we checked the efficiency of the six different extraction methods for intracellular metabolites as listed in Figure 1, in regard to metabolite recovery and reproducibility, and we determined the recovery of different classes of metabolites during the sample concentration step by using two mild techniques, lyophilization and reduced-pressure solvent evaporation.

Materials and methods

Strain and cultivation

The *Saccharomyces cerevisiae* strain CEN.PK113-7D (*MAT***a** *MAL2-8*^{c} *SUC2*) was used for the experiments. The strain was cultivated aerobically in shake flasks, using a rotary shaker at 150 r.p.m. and 30 °C. Each shake flask contained 150 ml medium composed of glucose (20 g/l), $(NH_4)_2SO_4$ (5.0 g/l), $MgSO_4 \cdot 7H_2O$ (0.5 g/l), KH_2PO_4 (3.0 g/l), vitamins and trace elements (Verduyn *et al*., 1992).

Chemicals

Methanol and chloroform were obtained from LAB SCAN Analytical Sciences (Dublin, Ireland). All

α-amino acids used were obtained from Serva Electrophoresis GmbH (Heidelberg, Germany) and all additional metabolite standards, as well as HEPES, PIPES, imidazole, methoxyamine, propidium iodide solution, and N-methyl-N-(trimethylsilyl)trifluoroacetamide, were purchased from Sigma (St. Louis, MO, USA). Perchloric acid, methyl chloroformate (MCF), sodium bicarbonate and sodium sulphate anhydrous were obtained from Merck (Darmstadt, Germany). EDTA was obtained from Fluka Chemie (Buchs, Switzerland). Pyridine, potassium hydroxide and sodium hydroxide were purchased from J.T. Baker (Deventer, Holland). All used chemicals were analytical grade.

Analytical procedures

The amino and non-amino organic acids and the nucleotides were determined simultaneously by GC–MS after MCF derivatization, according to the method described by Villas-Bôas *et al*. (2005). Sugars, sugar phosphates, sugar alcohols and peptides were determined simultaneously by GC–MS after methoxymation and silylation derivatization, following the protocol described by Roessner *et al*. (2000). The determination of extracellular sugar phosphates for the evaluation of leakage during quenching was performed using solid-phase extraction and high-pressure ion-exchange chromatography with pulsed amperometric detection (HPIC–PAD), according to Smits *et al*. (1998).

All GC–MS analyses were performed with a Hewlett-Packard system HP 6890 gas chromatograph coupled to a HP 5973 quadrupole mass selective detector, with electron ionization (EI) source operated at 70 eV. The capillary column used for all analysis was a J&W DB1701 (Folsom, CA, USA; 30 m × 250 μm i.d., 0.15 μm film thickness). The MS was always operated in scan mode (start after 5 min, mass range 38–550 *m/z* at 2.88 scans/s). Two different inlet inserts were used, depending on the derivatives to be analysed: Siltek gooseneck splitless deactivated liner (Restek, USA), 4.0 mm i.d., was used for analysis of MCF derivatives, and split straight liner with glasswool (Agilent Technologies, USA) for the TMS-derivatives. HPIC–PAD analysis was performed with a Dionex BioLC (Dionex, Sunnyvale, CA, USA) and the analytes were separated in a CarboPac PA1 column (Dionex) and a Dionex PA1 guard column at 30 °C, under conditions described by Smits *et al*. (1998).

The GC–MS data was analysed by the Automated Mass Spectral Deconvolution and Identification System (AMDIS; **http://chemdata.nist.gov/mass-spc/amdis/**) software, using a home-built library of MCF or TMS derivatives.

Membrane integrity assay

The membrane integrity of the yeast cells was assayed using propidium iodide (PI) according to the protocol described by Alfenore *et al*. (2003). PI is membrane-impermeant and can only penetrate the yeast cells if the membrane is damaged. PI binds to either DNA or RNA and forms a fluorescent complex.

The yeast cell pellets, after quenching, were resuspended in 1 ml saline solution (NaCl, 0.85 g/l) and then mixed with 0.5 ml propidium iodide solution (concentration 0.5 mg/ml; commercial solution, Sigma). After incubation for 10 min at 0 °C, the samples were inspected using fluorescence microscopy. The control samples consisted of yeast cells centrifuged and resuspended in saline solution without being in contact with methanol solution.

Quenching

Yeast cell cultures were quenched according to the method described by de Koning and van Dam (1992). Unless otherwise stated, the samples were quenched in triplicate at $OD_{600} = 6.0$ (middle exponential phase). Samples of culture (5 ml) were quickly harvested with a 5 ml automate Gilson pipette and released into 20 ml methanol–water solution (60% v/v) at −40 °C, resulting in a final methanol concentration of 50% (v/v) after quenching. The samples were carefully released into the centre of the methanol solution to avoid freezing on the sides of the tubes. The biomass was separated from the quenching solution by centrifugation at 770 × *g* for 20 min at −20 °C.

For evaluation of the quenching procedure, we quenched 5 and 10 ml samples released into 20 ml cold methanol solution at 60% (v/v) and 75% (v/v), respectively, in order to keep the same final methanol concentration of 50% (v/v). Both buffered (PIPES, 3 mM; EDTA, 3 mM; pH 7.0) and non-buffered cold methanol solutions

were tested. The temperature inside the quenching tubes during the quenching process was monitored using a digital thermometer (HI 955 501, Hanna Instruments, Italy). The biomass was separated from the quenching solution by centrifugation at 770 × *g* for 20 min at −20 °C according to the original protocol (de Koning and van Dam, 1992). A faster biomass separation from the quenching medium was also tested by centrifugation at 1540 × *g* for 5 min at −20 °C.

Extraction of intracellular metabolites

After quenching 5 ml yeast culture at exponential phase ($OD_{600} = 6.0$) using 60% (v/v) non-buffered cold methanol solution which, resulted in approximately 6 mg biomass, six methods for extraction of intracellular metabolites were evaluated by taking into account the number of compounds extracted, recovery of spiked standards and reproducibility. Figure 1 summarizes the methods evaluated and their respective protocols are described as follows:

- *Chloroform:methanol:buffer (CMB) extraction* (de Koning and van Dam, 1992). After quenching, the biomass was resuspended in 5.0 ml cold chloroform (−40 °C), followed by 2.5 ml cold methanol (−40 °C). Then 2.0 ml ice-cold buffer (PIPES, 3 mM; EDTA, 3 mM; pH 7.2) was added to the mixture under manual shaking. The suspension was agitated vigorously at 300 r.p.m. on a platform shaker for 45 min at −20 °C. Separation of the chloroform and the aqueous phase was induced by centrifugation (770 × *g*, 20 min, −20 °C). The upper aqueous phase, containing the polar metabolites, was collected without disturbing the pellet of the cell debris on top of the lower chloroform phase. A second extraction was initiated by adding 2.0 ml methanol (−40 °C) and 2.0 ml ice-cold buffer to the chloroform phase, followed by manual shaking for 30 s using a vortex mixer. After centrifugation (770 × *g*, 20 min, −20 °C) the upper aqueous phase was collected and pooled with the previous extract and the solvents were removed by lyophilization.
- *Boiling ethanol (BE) extraction* (Gonzalez *et al*., 1997). Boiling buffered ethanol solution (5 ml 75% v/v HEPES, 0.25 M, pH 7.5) was added to the cell pellet after quenching. The mixture was incubated for 3 min at 80 °C and immediately cooled in an ice-bath for 3 min. The cell residues were separated from the extract by centrifugation at 770 × *g* for 20 min at 4 °C. The supernatant was lyophilized.
- *Perchloric acid (PCA) extraction* (Hajjaj *et al*., 1998). After quenching, the biomass was resuspended in 6.4 ml ice-cold 10% (v/v) perchloric acid and 0.6 ml imidazole (1.0 M). The extraction was performed by three freeze–thaw cycles with vigorous shaking at 300 r.p.m. for 10 min between each cycle. After centrifugation for 5 min at 770 × *g* and 0 °C, the supernatant was neutralized with KOH (10 M) and the $KClO_4$ precipitate was removed by centrifugation. The supernatant was lyophilized.
- *Potassium hydroxide (KOH) extraction* (Hajjaj *et al*., 1998). After quenching the biomass was resuspended in 3.0 ml KOH (0.25 M) and the mixture was incubated at room temperature for 15 min. The cell residues were separated by centrifugation at 770 × *g* for 8 min (4 °C) and the supernatant of the samples to be silylated was neutralized with 10% (v/v) $HClO_4$ prior to lyophilization.
- *Methanol:water (MW) extraction* (Maharjan and Ferenci, 2003). The biomass was resuspended in 2.5 ml 50% (v/v) cold methanol solution (−40 °C) after quenching and the mixture was frozen in liquid nitrogen. The frozen suspension was then thawed in an ice-bath and centrifuged at 770 × *g* for 20 min at −20 °C. The supernatant was collected and an additional 2.5 ml of cold methanol solution 50% (v/v) (−40 °C) was added to the pellet and shaken manually for 30 s using a vortex mixer. The mixture was centrifuged at 770 × *g* for 20 min at −20 °C and both supernatants were pooled and lyophilized.
- *Pure methanol (PM) extraction.* This method was a slight modification of the previous method described by Maharjan and Ferenci (2003). The same procedure was applied, but the cold methanol solution 50% (v/v) was replaced by pure cold methanol (−40 °C).

Metabolite standards spiked in the samples

Representative compounds from different polar metabolite classes were selected to be spiked into the samples at the beginning of the different extraction procedures. The following metabolites (and their respective classes) were used: valine

(neutral amino acid); glutamate (acidic amino acid); lysine (basic amino acid); phenylalanine and tryptophan (aromatic amino acids); pyruvate and lactate (monocarboxylic acids); fumarate, succinate (dicarboxylic acids); citrate and isocitrate (tricarboxylic acids); 2-oxoglutarate and 2-oxoadipate (α-keto acids); myristate (fatty acid); NADH and NADP (nucleotides); glutathione (peptide); trehalose and xylose (sugars); glucose 6-phosphate, fructose 6-phosphate, mannose 6-phosphate, and ribose 5-phosphate (sugar phosphates); glycerol, mannitol, arabitol, and xylitol (sugar alcohols). Two metabolite mixture solutions were prepared: (a) 100 µl stock solutions (1 M) of each amino and non-amino organic acids and nucleotides were mixed and diluted to a final concentration of 10 µM/ml; (b) 400 µl each stock solution (0.5 M) of sugars, sugar derivatives and peptide were mixed and diluted to prepare a metabolite mixture presenting a final concentration of 20 µM/ml. Before extraction triplicate yeast biomass samples were spiked with 100 µl of metabolite mixture (a) or (b).

Sample concentration

The sample extracts were concentrated by lyophilization at low temperature (−56 °C) using a Christ Alpha 1–4 freeze-dryer (Osterode, Germany). An additional 10 ml MilliQ water was added to those extracts containing organic solvents in order to keep the samples frozen during the lyophilization process (H_2O content >75% (v/v)). The dried samples were stored for 7 days at −20 °C prior to the GC–MS analysis.

Evaluation of lyophilization and solvent evaporation

Three variables during lyophilization process were investigated: (1) deactivation of the lyophilization glass flask by silylation; (2) breakage of the vacuum after the process using an inert gas (N_2) or air; and (3) storage of the dried samples before analysis under reducing and dry atmosphere (N_2) at −20 °C or in ordinary freezer at −20 °C. A factorial optimization experiment was designed, as shown in Figure 2. All experiments were performed in triplicate and the test samples were prepared adding either 100 µl metabolite mixture (a) or (b) (see metabolite standards) to 10 ml MilliQ water.

Solvent evaporation was performed under vacuum using a Thermo Savant SC210A SpeedVac® Plus without heating for approximately 3.5 h. The test samples were prepared by adding 100 µl metabolite mixture (a) or (b) (see metabolite standards) to 3 ml pure methanol.

Variables	−	+
Flask deactivation	Non-silanized	Silanized
Vacuum breakage	Air	Nitrogen
Storage	–20°C, freezer	–20°C, dry and reducing atmosphere

Experiments	Flask deactivation	Vacuum breakage	Storage
I	–	–	–
II	–	–	+
III	–	+	+
IV	+	+	+
V	+	+	–
VI	+	–	–

Figure 2. Factorial optimization of the lyophilization process, using a reduced 2^3-factorial design with six instead of eight different experiments. Each experiment was performed in triplicate

Calculations

All calculations are performed using the amount estimated by AMDIS, which is equivalent to the peak area obtained by integration of the peak in the chromatogram.

Leakage of intracellular metabolites

The concentration of intracellular metabolites was estimated based on data previously reported by Villas-Boas *et al*. (2005) and the metabolite concentrations found in the quenching solution corresponding to similar quenching conditions used by Villas-Boas *et al*. (2005):

$$\hat{c}_{\text{intra}} = \frac{x_{\text{quenching}} + x_{\text{intra}}}{V_{\text{sample}}} - c_{\text{extra}} \qquad (1)$$

where $\hat{c}_{\text{intra}}$ is the estimated concentration of intracellular metabolites, $x_{\text{quenching}}$ is the measured total amount of metabolites in the quenching solution, x_{intra} is the determined amount of intracellular metabolites after quenching as reported by Villas-Boas *et al*. (2005), c_{extra} is the concentration of extracellular metabolites detected in the culture filtrate, and V_{sample} is the volume of the sample.

Therefore, the relative leakage of intracellular metabolites was calculated for each quenching method tested according to equation (2):

$$Leakage = \frac{x_{\text{quenching}}/V_{\text{sample}}}{\hat{c}_{\text{intra}}} \cdot 100\% \qquad (2)$$

Recovery of spiked metabolites after extraction and sample concentration

The recovery was calculated based on the total signal of the GC peaks. Values obtained from the extracted/concentrated samples were divided by the values obtained in the control samples according to equation (3):

$$Recovery = \frac{x_{\text{extract}}}{x_{\text{control}}} \cdot 100\% \qquad (3)$$

where x_{extract} is the total signal of the metabolite detected in the sample extract, and x_{control} is the total signal of the metabolite detected in the control sample (standard).

The control sample was 100 µl metabolite mixture (a) or (b) (see metabolite standards), directly lyophilized and analysed by GC–MS after proper derivatization. When the recovery results were higher than 100% (spiked + cell metabolites), the recovery was presented as >100%, which apply only for the evaluation of the extraction procedures. Recoveries above 100% were typically observed when high metabolite concentrations were extracted from the biomass.

Results

Quenching

Fast transfer of 5 ml fermentation broth at 30 °C into 20 ml 60% (v/v) methanol solution at −40 °C resulted in a cell suspension at −20 °C, returning to −40 °C within 3 min. Fast sampling of 10 ml fermentation broth at 30 °C into 20 ml 75% (v/v) methanol solution at −40 °C resulted in a cell suspension at −10 °C, re-establishing the original temperature (−40 °C) in 5 min.

The membrane integrity assay showed that the yeast cell membrane is damaged by contact with both buffered and non-buffered cold methanol solution at either 60% (v/v) or 75% (v/v) (Figure 3). The metabolite levels in the supernatant of quenched samples confirmed that the methanol solution caused damage to the yeast cell membrane (Table 1). Several organic and amino acids were practically extracted from the cells during the quenching process. On the other hand, sugar phosphates did not appear to leak from the intracellular medium because their levels were roughly the same as those found in the culture media (data not shown).

Increasing the speed and decreasing the time of centrifugation to separate the yeast cells after

Table 1. Relative metabolite leakage levels after different quenching experiments using cold methanol solution

	Experiments				
Metabolites	**I**	**II**	**III**	**IV**	**V**
Lactate	13	0	8	3	78
Fumarate	0	0	0	0	0
Malate	0	0	0	0	0
Succinate	0	0	0	0	0
Citramalate	0	0	92	0	~100
PEP[1]	~100	~100	~100	~100	0
cis-Aconitate[1]	~100	~100	~100	~100	0
Citrate	34	86	~100	51	0
Myristate	0	0	0	0	~100
Alanine	0	0	40	37	0
Glycine	8	10	16	14	12
Valine	76	56	86	58	8
Leucine	78	57	~100	59	32
Isoleucine	83	60	~100	61	6
Proline	~100	82	~100	83	8
Threonine	21	47	49	44	0
Aspartate	71	69	77	52	0
Phenylalanine	0	0	0	0	0
α-Ketoglutarate	0	0	0	0	0
Butyrate	70	77	~100	80	0
Glyoxylate	0	0	40	37	0
Asparagine	14	44	73	0	0
Glutamate	~100	71	~100	62	0
Methionine	0	81	0	0	0
Serine	0	79	0	0	0
Histidine	0	0	0	16	0
Ornithine	0	5	9	7	0
Lysine	6	18	26	24	0
Tyrosine	0	21	82	58	0

Values expressed in percentage of the total intracellular level. The total intracellular level is calculated from intracellular concentration (Villas-Bôas *et al.*, 2005) plus leaked levels (for details, see Materials and methods). Experiments: I, 5 ml sample to 20 ml 60% (v/v) cold methanol solution; II, 10 ml sample to 20 ml 75% (v/v) cold methanol solution; III, 5 ml sample to 20 ml 60% (v/v) buffered cold methanol solution; IV, 10 ml sample to 20 ml 75% (v/v) buffered cold methanol solution; V, 5 ml sample to 20 ml 60% (v/v) cold methanol solution and biomass separation under fast centrifugation (5 min at 1540 × *g*).
[1] Metabolites not detected intracellularly.

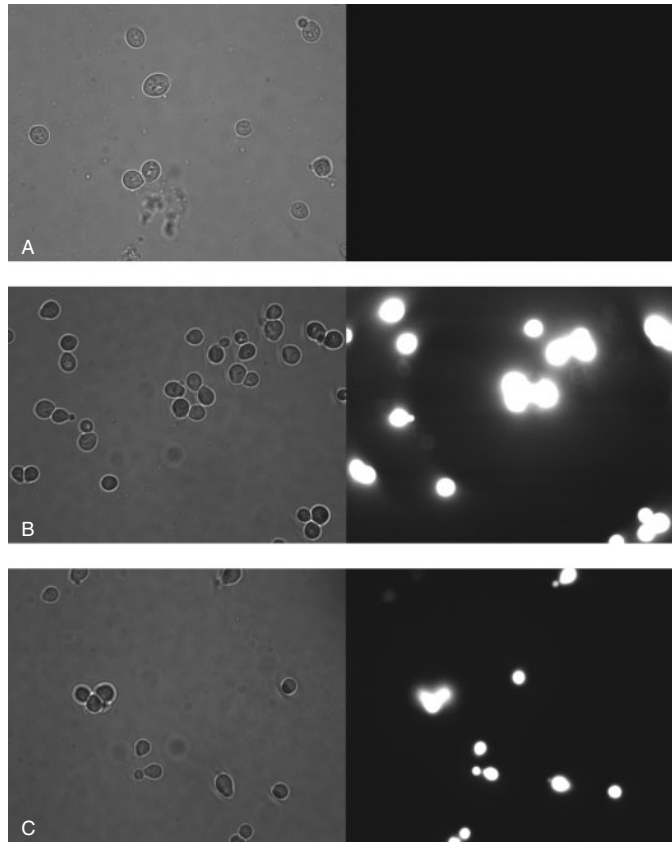

Figure 3. Yeast cells assayed with propidium iodide to test the membrane integrity during quenching with cold methanol solution at −40 °C. (A) Control cells centrifuged and resuspended in saline solution. (B) Cells quenched with 60% (v/v) methanol solution at −40 °C. (C) Cells quenched with 60% (v/v) buffered methanol solution at −40 °C

quenching resulted in a considerably reduced number of metabolites that leaked from the cells (Table 1), showing that the longer the time the cells stay in contact with the quenching solution, the higher is the leakage of intracellular metabolites, which is especially the case for amino acids. However, a few metabolites presented higher leakage when compared to the original protocol (de Koning and van Dam, 1992), e.g. lactate, citramalate and myristate were detected at a higher level in the supernatant of samples centrifuged at higher speed (Table 1).

Intracellular metabolite extraction

None of the extraction methods tested ensured a 100% recovery of all classes of metabolites spiked into the samples. Each method favoured some classes of compounds and even some discrimination within the same class was observed. Table 2 presents the global recovery for the different classes of metabolites: that is the average value of the recovery of different representatives for each class (some classes have only one representative).

In particular, the acidic amino acid glutamate showed 100% recovery in all methods tested.

Table 2. Global recovery of the different metabolite classes after different extraction method ($n = 3$)

Metabolite class	CMB (%)	BE (%)	PCA (%)	KOH (%)	MW (%)	PM (%)
Amino acid (neutral)	>100	87	9	96	58	96
Amino acid (basic)	100	36	0	0	36	>100
Amino acid (acidic)	>100	>100	>100	>100	>100	>100
Amino acid (aromatic)	100	49	2	>100	75	85
Organic acid (1-COOH)	75	90	2	13	>100	98
Organic acid (2-COOH)	77	97	6	72	98	94
Organic acid (3-COOH)	85	66	0	3	60	72
Organic acid (α-keto)	53	58	1	3	49	82
Fatty acid	0	43	5	64	50	71
Nucleotide	54	3	8	22	62	83
Peptide	57	31	49	41	28	44
Sugar	56	2	54	21	27	30
Sugar alcohol	69	50	68	80	69	78
Sugar phosphate	27	0	0	0	0	0

Extraction method: CMB, chloroform : methanol : buffer; BE, boiling buffered ethanol; PCA, perchloric acid; KOH, potassium hydroxide; MW, cold methanol solution 50% (v/v); PM, pure cold methanol.

Table 3. General classification of the different extraction methods based on the recovery of spiked metabolite standards ($n = 3$)

	Classes of metabolites							
Method	**Amino acids**	**Organic acids**	**Fatty acids**	**Nucleotides**	**Peptides**	**Sugars**	**Sugar alcohols**	**Sugar phosphates**
CMB	*****	****	nr	***	***	***	****	**
BE	****	****	***	*	**	*	***	nr
PCA	**	*	*	*	***	***	****	nr
KOH	****	**	****	**	***	**	****	nr
MW	****	****	***	****	**	**	****	nr
PM	*****	*****	****	*****	***	**	****	nr

Extraction method: CMB, chloroform : methanol : buffer; BE, boiling buffered ethanol; PCA, perchloric acid; KOH, potassium hydroxide; MW, cold methanol solution 50% (v/v); PM, pure cold methanol.
***** >80%; **** >60%; *** >40%; ** >20%; * >0%; nr, not recovered.

However, the high recovery of glutamate can be explained by the high natural intracellular concentration of this metabolite in *S. cerevisiae* (Villas-Bôas *et al*., 2005). All methods recovered more than 100% glutamate, which makes it difficult to define how much was lost. In contrast, the sugar phosphates were recovered only by the CMB method. NADH and NADP showed very distinct recoveries, resulting in bigger standard deviations, since NADP was much better recovered than NADH (data not shown). The recovery of the fatty acid, peptide, sugars and sugar phosphates did not reach 80% in any extraction method tested.

Table 3 rates the different methods according to their capability in recovering the spiked classes of metabolites. The CMB method presented excellent recoveries of amino acids, organic acids and sugar alcohols, and was the only method able to recover part of the sugar phosphates spiked into the samples (Table 2). However, the nucleotides were poorly recovered and the fatty acid was not detected at all.

Extraction with BE resulted in reasonable recoveries of amino and non-amino organic acids including the fatty acid (Table 3). This method especially recovered the mono- and dicarboxylic acids better. Nonetheless, sugars, sugar phosphates and nucleotides were poorly recovered and the peptide, tricarboxylic acids and basic and aromatic amino acids showed comparatively poorer recoveries (Table 2).

Acidic extraction with PCA presented the worst recoveries in general and the extracts after lyophili-

zation contained a large amount of salt ($KClO_4$), making it difficult to redissolve for derivatization. Only sugars, sugar alcohols and the peptide presented relatively good recoveries. All the other metabolite classes presented recoveries lower than 10% (except glutamate). Similarly, alkaline extraction using KOH also resulted in a large amount of salt after the sample concentration step, giving the same analytical problems. However, the latter presented good recoveries of amino acids (except basic amino acid), fatty acid and sugar alcohols (Table 2). Nucleotides, sugars, sugar phosphates and most of the organic acids were poorly recovered by extraction with KOH.

Although not presenting excellent recoveries of neutral and basic amino acids, sugars and sugar phosphates, the MW extraction recovered most of metabolite classes reasonably well. Particularly the nucleotides were significantly better recovered by this method. However, by changing the methanol concentration from 50% (v/v) to pure cold methanol (PM) it was possible to get outstanding metabolite recoveries for most of the metabolite classes tested (Table 3). The modified method presented the best recovery results for nucleotides and α-keto acids. Only sugars and sugar phosphates showed relative poor recoveries (Table 2). Furthermore, the pure methanol method achieved the best reproducibility compared to the other methods.

Figure 4 shows the global extraction results based on the number of peaks detected and identified by GC–MS after MCF derivatization.

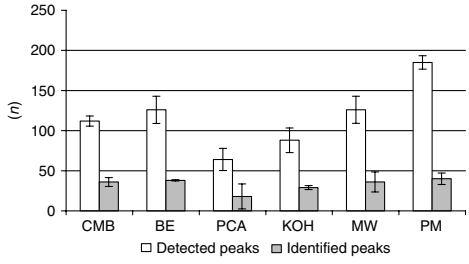

Figure 4. Total number of MCF derivatives detected and identified by GC–MS after different extraction procedures ($n = 3$). CMB, chloroform : methanol : buffer; BE, boiling buffered ethanol; PCA, perchloric acid; KOH, potassium hydroxide; MW, cold methanol solution 50% (v/v); PM, pure cold methanol

Although not every GC peak corresponds to a single metabolite, because one metabolite can yield more than one product after the derivatization reaction, the number of detected peaks and the number of identified metabolites is undoubtedly a strong indication of the extraction efficiency. Again PM extraction yielded the highest number of GC peaks (>150) as well as the highest number of identified peaks. Besides the highest number of metabolite extracted, this method presented an excellent consistency in the number of compounds extracted (Figure 4). CMB, BE and MW extractions yielded a quite similar number of GC peaks (fewer than 100) and identified compounds. However, among them, CMB gave the best reproducibility. Extractions at extreme pH, alkaline and particularly acidic extractions presented the poorest efficiency with respect to extraction of metabolites. Extraction with PCA resulted in the lowest number of GC peaks and identified compounds as well as this method had the lowest reproducibility (Figure 4).

Sample concentration

The lyophilization process resulted in a very diverse recovery pattern depending on the class of metabolites, and similarly to the extraction results, discrimination was observed even within the same metabolite class. Table 4 presents the overall recovery of the different metabolite classes after the lyophilization process. Basically, all organic acids, nucleotides, sugars and the peptide presented recovery higher than 80% under one or more lyophilization condition tested. The fatty acid was poorly recovered in all conditions and the basic amino acid lysine was practically not recovered at all. The recovery of sugar alcohols and sugar phosphates was <75%.

The experiment was designed as a reduced 2^3-factor experiment and was evaluated by analysis of variance (ANOVA). The deactivation of the lyophilization flasks (50 ml Erlenmeyer) by silanization was the only variable that slightly positively affected the recovery of metabolites, but the only affected classes were the aromatic amino acids and nucleotides. Keeping the lyophilized samples in a dry and reducing environment slightly improved the recovery of sugars and, interestingly, the breakage of vacuum with nitrogen seems to affect negatively the recovery of the fatty acid, sugar alcohols and sugar phosphates. Overall, the

Table 4. Global recovery of the different metabolite classes after lyophilization under different conditions ($n = 3$)

	Lyophilization conditions					
Class of metabolites	**I**	**II**	**III**	**IV**	**V**	**VI**
Amino acid (neutral)	95	92	92	99	98	99
Amino acid (basic)	0	0	0	3	5	0
Amino acid (acidic)	100	83	100	100	100	100
Amino acid (aromatic)	50	41	52	59	54	44
Organic acid (1-COOH)	100	94	96	100	100	96
Organic acid (2-COOH)	100	96	91	100	100	96
Organic acid (3-COOH)	100	92	100	91	100	98
Organic acid (α-keto)	90	76	63	86	81	85
Fatty acid	42	24	9	32	22	47
Nucleotide	48	38	59	70	84	71
Peptide	100	87	87	66	75	91
Sugar	66	84	90	86	80	70
Sugar alcohol	73	68	67	49	61	71
Sugar phosphate	74	68	46	46	52	71

Variables: FD, flask deactivated by silanization; FND, flask not deactivated; VO_2, vacuum broken with air; VN_2, vacuum broken with nitrogen; SF, samples stored after being dried in freezer at $-20\,^{\circ}C$; SDN_2, samples stored after being dried in dry and reducing atmosphere at $-20\,^{\circ}C$.
Conditions: I, FND/VO_2/SF; II, FND/VO_2/SDN_2; III, FND/VN_2/SDN_2; IV, FD/VN_2/SDN_2; V, FD/VN_2/SF; VI, FD/VO_2/SF.

three variables investigated did not significantly change the recovery of the different metabolite classes after lyophilization.

On the other hand, concentration of samples by solvent evaporation under vacuum presented an excellent recovery for all amino acids, organic acids, nucleotides, sugars and sugar alcohols (Table 5). About 100% of amino acids, nucleotides and organic acids (excluding the fatty acid) were recovered after solvent evaporation. Only sugar phosphates and glutathione (peptide) presented a comparatively lower recovery, 67% and 56%, respectively. However, this procedure is only suitable for extraction methods based on organic solvents if there should be a fast sample concentration. Aqueous extracts would take a considerably longer time for concentration and use of heating would be required.

Table 5. Global recovery of different metabolites after solvent evaporation under reduced pressure at room temperature ($n = 3$) RSD, relative standard deviation

Metabolites	**Recovery (%)**	**RSD ($n = 3$)**
Valine	100	3.5
Lysine	100	10.7
Glutamate	100	8.8
Phenylalanine	100	6.8
Tryptophane	100	8.7
Pyruvate	100	10.0
Lactate	100	6.0
Fumarate	100	9.3
Succinate	100	8.9
Citrate	100	8.7
Isocitrate	100	10.0
2-Oxoglutarate	100	13.2
2-Oxoadipate	100	11.7
Pimelate	100	11.9
Myristate	83	4.3
NADP	100	10.5
NADH	100	11.6
Glutathione	56	11.0
Xylose	85	10.0
Trehalose	100	1.0
Glycerol	92	10.0
Xylitol	85	3.0
Arabitol	85	3.0
Mannitol	62	8.0
Ribose 5-phosphate	42	4.0
Glucose 6-phosphate	100	8.0
Fructose 6-phosphate	84	10.0
Mannose 6-phosphate	41	4.0

Discussion

The results from quenching yeast cultures with cold methanol solution were surprising. This is the most frequently employed method for quenching microbial cells and several previous works have reported the absence of significant intracellular metabolite leakage using this method (de Koning and van Dam, 1992; Gonzalez *et al*.,1997; Hans *et al*., 2001). However, these results explain the high variability of microbial metabolome data as well as the discrepancies often found between laboratories, usually attributed to differences in strain backgrounds and/or bad reproducibility of different cultivations.

Although Hans *et al*. (2001) did not detect significant leakage of amino acids during quenching using cold methanol solution, our results show that several amino acids leak significantly from the yeast cells after contact with cold methanol solution. However, these authors applied faster biomass separation from the quenching solution, using higher centrifugation speed for a shorter time (5 min), and they buffered their quenching solution at pH 7.5. Therefore, their report is actually in agreement with our results, because we only observed a minor leakage of a few amino acids

(mainly leucine and glycine) under very similar conditions (Table 1, experiment V).

On the other hand, a significant leakage of organic acids from the yeast cells, mainly phosphoenolpyruvate and *cis*-aconitate, was observed following the original protocol proposed by de Koning and van Dam (1992). However, we did not detect leakage of pyruvate, which is in agreement with the original results reported by de Koning and van Dam (1992). Similarly, we also confirmed the results reported by Gonzales *et al*. (1997), where detectable leakage of sugar phosphates and nucleotides were not observed through quenching yeast cells in cold methanol solution.

Yeast cells, like bacterial cells (Jensen *et al*., 1999; Letisse and Lindley, 2000; Wittmann *et al*., 2004), are therefore concluded to be sensitive to cold methanol solution either buffered or non-buffered (Figure 3). Even varying the concentration of cold methanol solution from 25% to 75% did not prevent cell membrane damage as assayed with propidium iodide (data not shown). In addition, we did not observe sample freezing in any quenching condition tested, which could be pointed out as a reason to provoke membrane damage. However, by decreasing the time in which the yeast cells were exposed to cold methanol solution, the leakage of intracellular metabolites was generally minimized.

Furthermore, the possibility of quenching larger sample volumes into a smaller volume of cold quenching solution would be beneficial for many laboratories in order to achieve easier automation, as well as obtaining a higher final biomass content per sample. The dilution of yeast fermentation broths very often imposes severe limitations on the detection of intracellular metabolites. Doubling the volume of sampling did not result in a drastic increase of the final mixture temperature ($-10\,^{\circ}C$), which was reestablished to $-40\,^{\circ}C$ within 5 min; however, it is recommended to keep the temperature below $-20\,^{\circ}C$ until extraction (Lange *et al*., 2001).

The leakage observed after quenching yeast cells with cold methanol solution does not exclude the use of this method in several approaches of metabolome analysis. However, precautions must be taken in order to minimize losses of intracellular metabolites. The longer the cells are in contact with the quenching solution, the higher the leakage will be; thus, the common practice of washing the cell pellet with cold methanol solution to eliminate the interference of extracellular metabolites should be reconsidered and probably avoided. Alternatively, Wittmann *et al*. (2004) proposed a protocol for fast separation of bacterial cells from extracellular media by fast filtration under vacuum and washing the biomass with four volumes of cold saline solution at room temperature (the whole filtration step including the washing was finished in less than 45 s). This method permitted authentic quantification of intracellular amino acid pools, which could probably be adapted to yeast cells. However, this method seems not to be suitable for precise analysis of metabolites with rapid turnover, e.g. phosphorylated intermediates.

Another important issue in sample preparation for metabolome analysis is the procedure for an efficient extraction of intracellular metabolites. Most of the extraction protocols currently being applied for extraction of intracellular metabolites of yeasts are protocols originally developed for the analysis of targeted compounds, and it is quite clear that they do not suit to the needs of many metabolome approaches where a global metabolite profile is desired.

By examining the different extraction procedures it has become evident that there is a strong influence of the extraction method on the metabolite profile of yeast cells. Similar conclusions were also achieved for filamentous fungi (Hajjaj *et al*., 1998) and bacterial cells (Maharjan and Ferenci, 2003). The recent methodologies developed to acquire accurate information on microbial metabolite levels (Buchholz *et al*., 2002; Soga *et al*., 2002; van Dam *et al*., 2002; Mashego *et al*., 2004; Villas-Bôas *et al*., 2005; and others) urgently need an extraction protocol that is able to trap a higher diversity of intracellular metabolites of yeasts. However, the yeast metabolome comprises a variety of compounds with a wide range of physical and chemical properties that make this task virtually impossible.

Within this context, and based on the results presented here, we can draw a hierarchical scale for the six methods evaluated here, concerning their discriminative power. Starting from the less discriminative to the most discriminative, PM < MW < CMB < BE < KOH < PCA (Table 3). It was evident that acidic and alkaline extractions do not suit the requirements for a global metabolome analysis. These methods could be very efficient in targeted analysis of metabolites, e.g. amines

(Bouchereau *et al*., 2000), and the data acquisition method should be carefully chosen in order to eliminate the interferences of high salt concentrations. According to Maharjan and Ferenci (2003), destruction of compounds such as pyruvate, nucleotides and phosphorylated sugars under extreme acidic or alkaline pH is well documented. In addition, the neutralization of the extracts may increase the losses of metabolites by absorbing to the $KClO_4$ precipitate.

On the other hand, the solubility of the metabolites in the extraction solution is related to their polarity. Taking the principle that 'like dissolves like', more polar extraction media will dissolve more polar compounds and vice versa, explaining mainly the losses observed here. For instance, the poor recovery of fatty acid during CMB extraction could have been resulted from it being dissolved in the chloroform phase, which was not analysed. Similarly, sugars, and sugar phosphates were badly recovered in most of the methods tested probably due to the lack of polarity to extract them efficiently. When the polarity was high, such as in PCA and KOH methods, these compounds were probably not stable at the extreme pH conditions applied. Sugar phosphates presented the worst recovery results and this can be attributed mainly to their instability. These metabolites are extremely labile, which makes their detection by GC–MS difficult. Therefore, the recovery results concerning to this particular metabolite class could benefit from using a more targeted sample work-up and a more appropriate analytical method than GC–MS (see Concluding remarks).

Loss of several metabolites were also observed during BE extraction; most likely because many metabolites are heat-labile. The sugars and the nucleotides in particular presented the worst recoveries, but even several amino and organic acids presented bad recoveries. Maharjan and Ferenci (2003) obtained similar results using BE to extract intracellular metabolites of bacteria. We excluded the evaporation step as proposed in the original BE protocol. Instead of evaporating the buffered ethanol solution after the extraction and resuspending the cell debris in water and then lyophilizing the aqueous extract after centrifugation, we lyophilized the buffered ethanol solution directly after centrifugation. It is possible that the spiked metabolites would have had better solubility in water rather than in buffered ethanol solution, resulting in better recovery values. Nonetheless, we argue that inclusion of the evaporation step would not have significantly changed the results presented here, since we got excellent recoveries using pure methanol (PM protocol) that has solubility properties similar to ethanol.

On the other hand, the extraction method proposed by Maharjan and Ferenci (2003) for intracellular metabolite extraction of bacteria presented distinct advantages over the other methods and showed to be well applicable to yeast cells (Tables 2 and 3). The modification of this method proposed in this present work overcame most of our expectations. Extraction using pure cold methanol presented the lowest discrimination, coupled to a quite simple sample work-up and excellent recovery and reproducibility. In addition, this method makes use of a single organic solvent that is not as toxic as chloroform and can be easily removed from the sample by solvent evaporation. Despite its limitations in the extraction and recovery of sugar and sugar phosphates, this method has excellent potential to be employed routinely in high-throughput metabolome analysis of yeasts; furthermore, it presents a great potential for automation.

Nonetheless, it is evident that the lyophilization is partly responsible for the poor recovery of many metabolite classes when comparing the different extraction procedures. Most metabolite classes presented losses during lyophilization. Therefore, the results from extractions and lyophilization could be taken into account in order to get a more precise view regarding to the efficiency of the different extraction methods. This would not change the conclusions of this work, since the relative performance of the method would remain unchanged.

However, the losses during lyophilization are quite significant, as this sample concentration technique was supposed to be a gentle methodology to concentrate aqueous samples. The losses of the different metabolites could be related to discrimination during resuspension. The different metabolites present different solubilities in the reduced volume of resuspension solution, which is 200 µl 1% (w/v) NaOH for MCF derivatization and 200 µl deionized water for TMS derivatization, followed by a second freeze-drying to eliminate the water. Therefore, discrimination during dissolving these solutes in a very small volume of solvent is likely to happen. In addition, considering that most extraction procedures end up with approximately 5–10 ml

extracts, we were forced to use relatively large flasks for lyophilization. Thus, to dissolve the remaining salts after the concentration process into the 200 µl of solvent was definitely a challenge, contributing for the losses observed. However, the resuspension was in general reproducible. Increasing the solvent volume for resuspension to cover a larger surface area in the flask could improve the recovery, but this would also require a second concentration step.

Organic solvent evaporation showed to be a viable method for concentration of samples containing primary metabolites. The excellent recovery achieved by this technique with reliable reproducibility (Table 5) indicates that this methodology is the best alternative for sample concentration in metabolome analysis. However, this technique is dependent on the type of extraction procedure used, since this method is not suitable for aqueous sample extracts unless acetone is added to the mixture to accelerate the evaporation process. Still, solvent evaporation poses several advantages over the lyophilization process, since it is faster and seems less aggressive and less discriminative. Considering that we have obtained the best extraction results using cold pure methanol as the extractive agent, solvent evaporation is undoubtedly the method of choice to concentrate the extracted samples. By combining the PM extraction procedure with sample concentration by solvent evaporation, it should be possible to improve the throughput of sample preparation in metabolome analysis of yeast.

Concluding remarks

This work aimed to evaluate the relative overall performance of the different commonly used sample preparation methods in the metabolomic field, by fixing the analytical method (GC–MS) and the biological sample (yeast cells in exponential phase during aerobic growth on glucose). For instance, one can argue that GC–MS was not the most adequate method for the analysis of the different classes of metabolites considered in this study. However, our aim was not to analyse the different metabolites accurately, but analyse them precisely in a pool, under the influence of many other metabolites, to become as close as possible to the reality of metabolome analysis. The results should asses the relative performance of several sample preparation steps. Furthermore, it was clear from the different recoveries obtained within the same metabolite class that the chosen representative metabolites do not reflect precisely the real metabolome complexity; rather, it gave us an insight into the complex microscopic world of small molecules that comprise the whole metabolome. Hence, the greater the accuracy of the analysis of these molecules, the closer we will be to puzzling out the *in vivo* cell function at the molecular level.

Acknowledgements

We are very grateful to Mrs Kianoush K. Hansen for helpful technical support and Dr Geoff Lane (AgResearch — Grasslands) for agreeing to review our text. This work has been supported by the Danish Biotechnological Instrument Centre.

References

Adams MA, Chen Z, Landman P, Colmer TD. 1999. Simultaneous determination by capillary gas chromatography of organic acids, sugars, and sugar alcohols in plant tissue extracts as their trimethylsilyl derivatives. *Anal Biochem* **266**: 77–84.

Alfenore S, Délia ML, Strehaiano P. 2003. Use of ATP measurements by bioluminescence to quantify yeast's sensitivity against a killer toxin. *Anal Chim Acta* **495**: 217–224.

Bouchereau A, Guénot P, Larher F. 2000. Analysis of amines in plant materials. *J Chromatogr B* **747**: 49–67.

Buchholz A, Hurlebaus J, Wandrey C, Takors R. 2002. Metabolomics: quantification of intracellular metabolite dynamics. *Biomol Eng* **19**: 5–15.

Buziol S, Bashir I, Baumeister A, *et al.* 2002. New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale. *Biotechnol Bioeng* **80**: 632–636.

Castrillo JI, Hayes A, Mohammed S, Gaskell SJ, Oliver SG. 2003. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **62**: 929–937.

Chen Z, Landman P, Colmer TD, Adams MA. 1998. Simultaneous analysis of amino and organic acids in extracts of plant leaves as *tert*-butyldimethylsilyl derivatives by capillary gas chromatography. *Anal Biochem* **259**: 203–211.

Conneely A, Nugent A, O'Keeffe M. 2002. Use of solid phase extraction for the isolation and clean up of a derivatized furazolidone metabolite from animal tissues. *Analyst* **127**: 705–709.

Cordeiro C, Freire AP. 1996. Methylglyoxal assay in cells as 2-methylquinoxaline using 1,2-diaminobenzene as derivatizing reagent. *Anal Biochem* **234**: 221–224.

Cremin P, Donnelly DMX, Wolfender JL, Hostettmann K. 1995. Liquid chromatography–thermospray mass spectrometric analysis of sesquiterpenes of *Armillaria* (Eumycota: Basidiomycotina) species. *J Chromatogr A* **710**: 273–285.

de Koning W, van Dam K. 1992. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal Biochem* **204**: 118–123.

Entian KD, Zimmermann FK, Scheel I. 1977. A partial defect in carbon catabolite repression mutants of *Saccharomyces cerevisiae* with reduced hexose phosphorylation. *Mol Gen Genet* **156**: 99–105.

Folch J, Lees M, Stanley GH. 1957. A simple method for the isolation and purification of total lipids from animal tissue. *Biol Chem* **226**: 497–509.

Gonzalez B, François J, Renaud M. 1997. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* **13**: 1347–1356.

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* **22**: 245–252.

Hajjaj H, Blanc PJ, Goma G, François J. 1998. Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi. *FEMS Microbiol Lett* **164**: 195–200.

Hans MA, Heinzle E, Wittmann C. 2001. Quantification of intracellular amino acids in batch cultures of *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **56**: 776–779.

Jensen NB, Jokumsen KV, Villadsen J. 1999. Determination of the phosphorylated sugars of the Embden–Meyerhoff–Parnas pathway in *Lactococcus lactis* using a fast sampling technique and solid phase extraction. *Biotechnol Bioeng* **63**: 357–362.

Kell DB. 2004. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **7**: 296–307.

Kopka J, Ohlrogge JB, Jaworski JG. 1995. Analysis of *in vivo* levels of acyl-thioesters with gas chromatography/mass spectrometry of the butylamide derivative. *Anal Biochem* **224**: 51–60.

Lange HC, Eman M, van Zuijlen G, *et al*. 2001. Improved rapid sampling for *in vivo* kinetics of intracellular metabolites in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **75**: 406–415.

Le Belle JE, Harris NG, Williams SR, Bhakoo KK. 2002. A comparison of cell and tissue extraction techniques using high-resolution ^{1}H-NMR spectrometry. *NMR Biomed* **15**: 37–44.

Letisse F, Lindley ND. 2000. An intracellular metabolite quantification technique applicable to polysaccharide-producing bacteria. *Biotechnol Lett* **22**: 1673–1677.

Maharjan RP, Ferenci T. 2003. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Anal Biochem* **313**: 145–154.

Martins AM, Cordeiro CA, Ponces Freire AM. 2001. *In situ* analysis of methylglyoxal metabolism in *Saccharomyces cerevisiae*. *FEBS Lett* **499**: 41–44.

Mashego MR, Wu L, van Dam JC, *et al*. 2004. Miracle: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotech Bioeng* **85**: 620–628.

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L. 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *Plant J* **23**: 131–142.

Ruijter GJ, Visser J. 1996. Determination of intermediary metabolites in *Aspergillus niger*. *J Microbiol Methods* **25**: 295–302.

Schaefer U, Boos W, Takors R, Weuster-Botz D. 1999. Automated sampling device for monitoring intracellular metabolite dynamics. *Anal Biochem* **270**: 88–96.

Shryock JC, Rubio R, Berne RM. 1986. Extraction of adenine nucleotides from cultured endothelial cells. *Anal Biochem* **159**: 73–81.

Smits HP, Cohen A, Buttler T, Nielsen J, Olsson L. 1998. Cleanup and analysis of sugar phosphates in biological extracts by using solid-phase extraction and anion-exchange chromatography with pulsed amperometric detection. *Anal Biochem* **261**: 36–42.

Soga T, Ueno Y, Naraoka H, *et al*. 2002. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal Chem* **74**: 2233–2239.

Tweeddale H, Notley-Mcrobb L, Ferenci T. 1998. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ('metabolome') analysis. *J Bacteriol* **180**: 5109–5116.

van Dam JC, Eman MR, Frank J, *et al*. 2002. Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. *Anal Chim Acta* **460**: 209–218.

van der Greef J, van der Heijden R, Verheij ER. 2004. The role of mass spectrometry in systems biology: data processing and identification strategies in metabolomics. In *Advances in Mass Spectrometry*, vol 16, Ashcroft AE, Brenton G, Monaghan JJ (eds). Elsevier: Amsterdam; 145–165.

Vaseghi S, Baumeister A, Rizzi M, Reuss M. 1999. *In vivo* dynamics of pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metabolic Eng* **1**: 128–140.

Verduyn C, Postma E, Scheffers WA, van Dijken JP. 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on regulation of respiration and alcoholic fermentation. *Yeast* **8**: 501–517.

Villas-Bôas SG, Moxley JF, Åkesson M, Stephanopoulos G, Nielsen J. 2005. High-throughput metabolite state analysis: the missing link in integrated functional genomics. *Biochem J* **388**: 669–677.

Visser D, van Zuylen GA, van Dam JC, *et al*. 2002. Rapid sampling for analysis of *in vivo* kinetics using the BioScope: a system for continuous-pulse experiments. *Biotechnol Bioeng* **79**: 674–681.

Weuster-Botz D. 1997. Sampling tube device for monitoring intracellular metabolite dynamics. *Anal Biochem* **246**: 225–233.

Wittmann C, Krömer JO, Kiefer P, Binz T, Heinzle E. 2004. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Anal Biochem* **327**: 135–139.

Zaldivar J, Borges A, Johansson B, *et al*. 2002. Fermentation performance and intracellular metabolite patterns in laboratory and industrial xylose-fermenting *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **59**: 436–442.

Chapter 6

Metabolic footprinting of *Saccharomyces cerevisiae* grown in microtiter plates

6.1 Introduction

Shaken cultures are widely used for microbial cultivation (Buchs, 2001). The design and setup is simple and this makes it easy to handle multiple cultures at a time. With the trend of increased throughput for screening multiple strains there is a considerable interest in miniaturization of the fermentation processes too (Duetz, 2007). The reduction of cultivation volume eases the handling, and the use of microtiter plates (MTPs) enables handling of several cultivations in parallel (Duetz et al., 2000). This increases the throughput of experiments and allows conduction of many replicates within experiments. This is beneficial in comparative studies as an increased number of replicates substantially improve the statistics and hereby better allow unraveling differences between strains. Also the inter-sample variance within a MTP can be reduced compared to running separate shake flasks (Borner et al., 2007), and this will further strengthen the data analysis.

Duetz et al. (2000) performed a detailed evaluation of the performance of MTPs, by investigating these with respect to mass transfer using a biological and chemical method to quantify the maximum oxygen transfer rate. The MTP setup was suggested for handling of culture collections and genome libraries. The group of Sauer also has numerous examples of using a 96-well MTP setup to culture different microorganisms (Blank et al., 2005a,b; Fischer et al., 2004; Fischer and

Sauer, 2005; Sonderegger et al., 2005). In their studies the cultures are primarily used for metabolic flux analysis using ^{13}C-labeled glucose. This has enabled *in vivo* flux estimation of up to 137 mutants in one study (Fischer and Sauer, 2005) and thereby presenting a technique for phenotyping a large number of strains, which is complementing high-throughput approaches like microarrays and proteome analysis (Sauer, 2004). For metabolite profiling Allen et al. (2003) used a commercial 100-well MTP for growing yeast prior to metabolic footprinting. In this study the optical density was automatically measured to ensure that the cultures had reached the stationary phase at the time of sampling. Another example of metabolite profiling involved direct integration with GC-MS and here the cells were cultured, extracted and derivatized directly in the MTP (Borner et al., 2007).

The purpose of this study was to evaluate the feasibility of cultivation in MTPs rather than traditional Erlenmeyer shake flasks beyond the default carbon source glucose. We investigated the growth of *Saccharomyces cerevisiae* in MTP cultures with glucose, galactose or ethanol as the sole carbon source. Glucose is the most common carbon source for growing *S. cerevisiae* and the only one used in previous studies based on cultivation in MTPs. Galactose is a another fermentative sugar, which is less repressive compared to glucose and might be used for induction of genes expressed under the GAL promoters. Both glucose and galactose are fermentable sugars, thus oxygen is not required for growth. Ethanol is an example of a more reduced substrate that requires oxygen to be metabolized. In order to evaluate the cultures metabolic footprinting was used to analyze the effect of different carbon sources.

6.2 Materials and methods

6.2.1 Strain

The *Saccharomyces cerevisiae* strain BY4741 (*MAT***a** *his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0) obtained from the European Saccharomyces cerevisiae archive for functional analysis (EUROSCARF, Frankfurt, Germany) was used in the study (Brachmann et al., 1998).

6.2.2 Medium

The strain was cultivated in a minimal synthetic medium supplemented with a metabolite cocktail according to Allen et al. (2003). The metabolite cocktail contained a selection of amino acids, organic acids and organic bases. The medium was based on yeast nitrogen base (YNB) without amino acids (BD Difco, Franklin

Lakes, NJ, USA) and the metabolite cocktail was composed of L-arginine, L-aspartate, L-glutamate, L-histidine, L-leucine, L-lysine, L-methionine, L-serine, L-threonine, L-tryptophan, L-valine, citrate, fumarate, malate, pyruvate, succinate, cytosine and uracil. All compounds from the metabolite cocktail had a concentration of 1 mM in the final medium. Cultivations were performed with either glucose, 20 g L^{-1}; galactose, 20 g L^{-1}; or ethanol, 5 g L^{-1}; as carbon source.

6.2.3 Pre-culture

Pre-cultures were prepared by inoculating a colony in 5 mL synthetic minimal medium with metabolite cocktail and glucose as carbon source. The pre-cultures were grown for 12 hours in an incubator at 30°C and 150 rpm.

6.2.4 Shake flask cultivation

Shake flask cultivations were performed in 500-mL baffled flasks with 50 mL medium. The medium was inoculated with the pre-cultures to $OD_{600} = 0.01$ and incubated at 30°C and different shaking speeds.

6.2.5 Microtiter plate cultivations

Cultivations were carried out in MTPs as demonstrated by Duetz et al. (2000). A 96-well plate with squared 2-mL wells (Whatman, Middlesex, UK) was covered with a 4-mm thick plate of spongy silicone. The spongy silicone plate was perforated with 96 1.5-mm holes located exactly above the center of each well. The spongy silicone plate was covered with 2-mm cotton wool and finally covered by a rigid lid of 6-mm aluminium. The aluminium lid had 96 6-mm holes exactly positioned above each of the wells. The aluminium plate was tightened by two screws. For MTP cultures, a culture with $OD_{600} = 0.01$ was prepared from the pre-culture into fresh medium. Aliquots of 500 μL of this culture were added to the respective wells in the MTP. The cultures were incubated at 30°C on a rotary shaker (Kühner AG, Schweiz) at 400 rpm with a shaking diameter of 25 mm.

6.2.6 Growth rate determination

For the cultivations the specific growth rates were estimated by measuring optical density (OD) at 600 nm on a Hitachi U-1100 spectrophotometer (San Jose, CA). The OD_{600} was assumed to be proportional with cell dry weight for OD_{600} values below 0.3. Thus the samples were diluted if necessary to stay below an OD_{600} of 0.3.

6.2.7 Analysis of extracellular metabolites

The samples were centrifuged (6400 rpm, 15 min) to remove biomass and stored at $-20°$C. Before analysis 150 μL of each sample were mixed with 150 μL of methanol. Metabolic footprints were obtained by direct infusion mass spectrometry (MS). A Micromass Q-Tof (Manchester, UK) mass spectrometer was equipped with a NanoMate device from Advion Biosciences (Ithaca, NY) for analysis by nano-electrospray ionization MS. The source block temperature was held at 60°C. High-resolution continuum data was acquired in negative mode in the mass range of 40-500 Da for 2 min with a scan time of 1.0 s and interscan of 0.1 s. The capillary voltage was set to +1.43 kV and 0.09 psi head pressure. The cone voltage was 20 V.

6.2.8 Data analysis

The mass spectra were converted by in-house written software and normalized by the total sum of ion intensities. The spectra were aligned in bins of 0.1 Da widths between 40 and 500 Da. The data was organized into at matrix with each row representing a sample and the columns containing the variables corresponding to the binned masses from the mass spectrum. All variables (columns in the data matrix) only containing zeros were removed from the data. Before principal component analysis (PCA) the individual variables were centered and scaled to unit variance. Analysis of variance (ANOVA) was used to select medium specific variables. The data analysis was executed in MATLAB (MathWorks, Natick, MA, USA).

6.3 Results

Initially, the yeast was cultivated in shake flasks to get a reference of specific growth rates for comparison with the MTP cultures. The results from the shake flask experiment are summarized in Table 6.1. Three different shaking speeds were used for the cultures and it turned out to have a significant ($\alpha = 0.05$) influence on the measured specific growth rate on glucose. The specific growth rate varied from 0.357–0.420 h^{-1} for glucose and increased with increasing shaking speed. Specific growth rates of 0.306 h^{-1} and 0.150 h^{-1} were measured in shake flasks with respectively galactose and ethanol as carbon source. The influence of shaking effect was not tested for these two substrates.

To characterize the growth of *S. cerevisiae* in MTPs and to investigate the variation between wells the BY4741 strain was cultivated on glucose, galactose or ethanol in MTPs. OD measurements were performed to determine the spe-

Table 6.1 Specific growth rates obtained from shake flask cultivation ($n = 3$) of *Saccharomyces cerevisiae* BY4741 on different carbon sources at different shaking speeds

Carbon source	Shaking speed [rpm]	Specific growth rate [h^{-1}]
Glucose	150	0.420 ± 0.007
	133	0.401 ± 0.003
	113	0.357 ± 0.006
Ethanol	113	0.150 ± 0.005
Galactose	113	0.306 ± 0.004

Table 6.2 Specific growth rates observed during micro-cultivation of *Saccharomyces cerevisiae* BY4741 on glucose, galactose and ethanol

Carbon source	Specific growth rate [h^{-1}]	Correlation coefficient
Glucose	0.438	0.9987
Galactose	0.307	0.9961
Ethanol	0.145	0.9985

cific growth rate. At each time point OD was measured in two or three different wells. The whole content of a well was used for measuring OD, thus sequential measurements were never made from the same well. The results are presented in Figure 6.1, and Table 6.2 summarizes the growth characteristics. *S. cerevisiae* grew exponentially until the stationary phase on both glucose and galactose. However, during the cultivation of ethanol a sudden slow-down in growth rate was observed from exponential to approximately linear growth. Specific growth rates of 0.438 h^{-1}, 0.307 h^{-1} and 0.145 h^{-1} were observed in the exponential growth phase on glucose, galactose and ethanol, respectively.

Samples for metabolic footprinting were taken from the MTP cultures. Due to different specific growth rates the samples were taken after 24 hours for glucose and galactose cultures and after 64 hours for the ethanol culture to ensure that the stationary growth phase was reached. A principal component analysis (PCA) was performed for dimension reduction. The analysis was capable of capturing 64% of the total data variance within the first and second principal component.

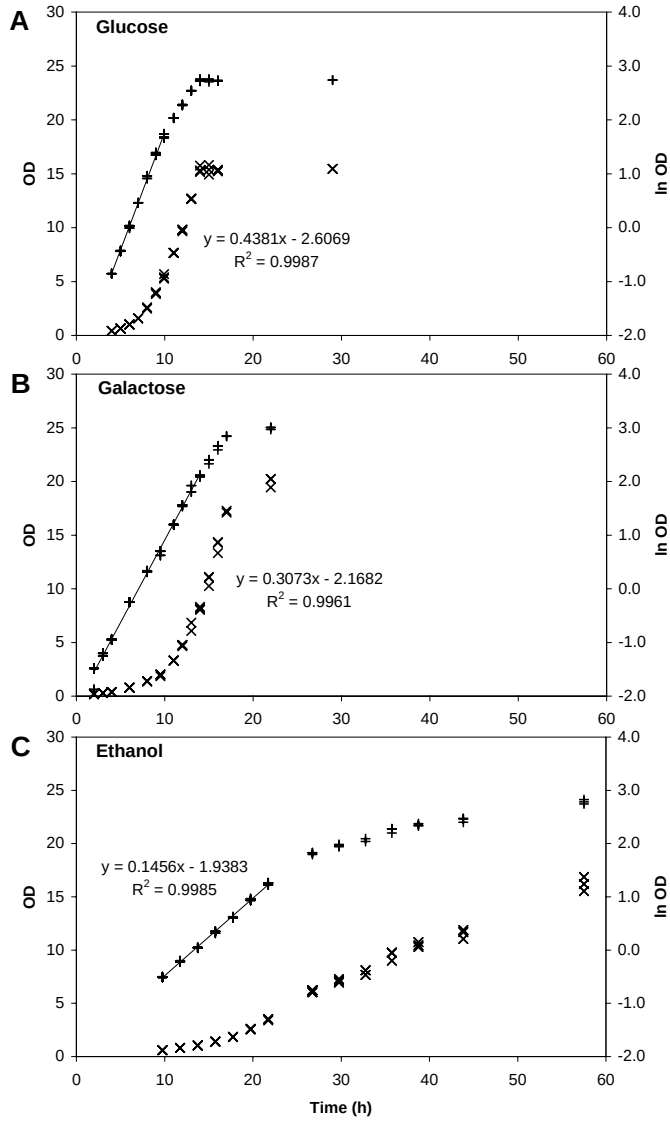

Figure 6.1 Growth of *S. cerevisiae* in microtiter plates on either A: glucose, B: galactose, or C: ethanol

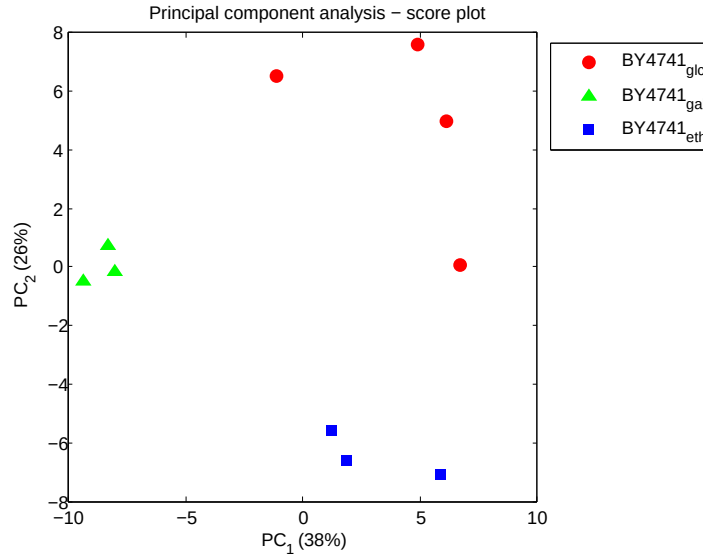

Figure 6.2 Principal component analysis (PCA) of metabolic footprints of *S. cerevisiae* BY4741 grown on glucose, galactose and ethanol. The PCA was based on binned and normalized mass spectra.

The samples were clearly clustering according to carbon source as seen in the score plot in Figure 6.2. The clusters of ethanol and galactose were well-defined, whereas the glucose samples were positioned further apart from each other and giving a more disperse cluster.

Statistical analysis of variance (ANOVA) was used to extract variables that were specific for the individual carbon sources. From ANOVA ions of 179.1 Da and 277.1 Da were found to be significantly higher for all samples taken from the cultures grown on galactose. This was most likely the deprotonated molecular ion $[M-H]^-$ and a hydrogen sulphate adduct $[M+HSO_4]^-$ of galactose indicating that the galactose was not fully depleted at the time of sampling. Another few ions were found to be specific for the different carbon sources, *e.g.* ion 147.0 and 189.0 for glucose, ion 175.4 and 177.1 for galactose, and ion 149.1 for ethanol.

6.4 Discussion

The specific growth rates observed in MTPs (Table 6.2) were very similar to those obtained in shake flasks (Table 6.1), but the specific growth rate on glucose for

the MTP culture was slightly higher than what was observed for shake flask cultivations. Taking into account the variation caused by shaking speed the results match very well. The observation of correlation between shaking speed and specific growth rate could be a coincidence, since different shakers were used and the temperature might not have been exactly the same.

The slow-down in growth rate on ethanol (Figure 6.1C) is most likely caused by oxygen limitation, which is set by the oxygen transfer rate. The maximum oxygen transfer rate is determined by the well geometry, shaking speed, shaking diameter, and culture volume as these factors affect the mixing and surface-volume ratio (Duetz et al., 2000; Duetz and Witholt, 2001, 2004; Duetz, 2007; Kensy et al., 2005). In these experiments the shaking diameter was 25 mm and increasing it to 50 mm would increase the maximum oxygen transfer rate by approximately three times (Duetz and Witholt, 2001, 2004), which would extend the exponential growth phase on ethanol and the stationary phase would be reached faster.

Looking into the metabolic footprinting obtained by negative direct infusion mass spectrometry, the individual samples clustered according to the carbon source. The differences in the culture media were only the carbon source. Substituting glucose with galactose has practically no impact on the direct infusion mass spectrum of the medium, since glucose and galactose have the same molecular composition and MS ionization. In the ethanol medium, the $C_6H_{12}O_6$ associated ions from glucose or galactose (*e.g.* $[M-H]^-$ and $[M+HSO_4]^-$) are absent in the mass spectrum and no particular ionization of ethanol is expected. Secondly, ethanol is known to be a product from respiro-fermentative growth of *S. cerevisiae* on glucose and galactose. Despite these minor medium differences footprinting of BY4741 proves to be specific in terms of physiology and differences in metabolism.

In order to explore genotypes, it is important to challenge the metabolic network by varying the primary carbon source for the experiments and this way finding fitness of growth (phenotype). Growing *S. cerevisiae* on different carbon sources is possible in MTPs as shown in this study, and the application of MTPs in connection with metabolic footprinting is an option as earlier demonstrated by Allen et al. (2003). Although Allen et al. (2003) argued that high cell densities was sufficient to inhibit growth on ethanol after the cultures had entered the stationary phase, taking the results of this study into account we have shown that *S. cerevisiae* grows well on ethanol in MTPs. As ethanol is a product of respiro-fermentative batch growth of *S. cerevisiae* on glucose, a significant amount of ethanol will be present in the medium after glucose in depleted; and therefore changes in the metabolic footprint will occur as ethanol is metabolized, but this is of cause dependent on the oxygen transfer rate.

The standardized format of MTPs calls for automation, which will allow high-

throughput screening. Even if a robot is not available, several tools can increase throughput like multi-channel pipettes, MTP spectrophotometers, special centrifuge holders *etc.* One concern of cultivation in MTPs is the difficulty of handling the individual wells in the plate independently. When different strains are inoculated in the different wells of a MTP, all wells should not be harvested at the same time due to differences in lag time and growth rates.

6.5 Conclusion

The scale-down from shake flask (50 mL) to MTP (500 μL) was successful. The cultivation in shaken MTPs mimics the growth in shake flasks very well for *S. cerevisiae* on different carbon sources. Cultivation in MTP has a major potential for handling several cultures in parallel, but for metabolic footprinting this will ideally require automation to monitor growth and allow independent sampling.

References

J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.*, 21(6):692–696, 2003.

L. M. Blank, L. Kuepfer, and U. Sauer. Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.*, 6(6):R49, 2005a.

L. M. Blank, F. Lehmbeck, and U. Sauer. Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res.*, 5:545–558, 2005b.

J. Borner, S. Buchinger, and D. Schomburg. A high-throughput method for microbial metabolome analysis using gas chromatography/mass spectrometry. *Anal. Biochem.*, 367:143–151, 2007.

C. B. Brachmann, A. Davies, G. J. Cost, E. Caputo, J. Li, P. Hieter, and J. D. Boeke. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, 14(2):115–132, 1998.

J. Buchs. Introduction to advantages and problems of shaken cultures. *Biochem. Eng. J.*, 7(2):91–98, 2001.

W. A. Duetz. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. *Trends Microbiol.*, 15(10):469–475, 2007.

W. A. Duetz and B. Witholt. Effectiveness of orbital shaking for the aeration of suspended bacterial cultures in square-deepwell microtiter plates. 7(2):113–115, 2001.

W. A. Duetz and B. Witholt. Oxygen transfer by orbital shaking of square vessels and deepwell microtiter plates of various dimensions. *Biochem. Eng. J.*, 17:181–185, 2004.

W. A. Duetz, L. Ruedi, R. Hermann, K. O'Connor, J. Buchs, and B. Witholt. Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates. *Appl. Environ. Microbiol.*, 66(6):2641–2646, 2000.

E. Fischer and U. Sauer. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat. Genet.*, 37(6):636–640, 2005.

E. Fischer, N. Zamboni, and U. Sauer. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. *Anal. Biochem.*, 325(2):308–316, 2004.

F. Kensy, H. F. Zimmermann, I. Knabben, T. Anderlei, H. Trauthwein, U. Dingerdissen, and J. Buchs. Oxygen transfer phenomena in 48-well microtiter plates: Determination by optical monitoring of sulfite oxidation and verification by real-time measurement during microbial growth. *Biotechnol. Bioeng.*, 89(6):698–708, 2005.

U. Sauer. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.*, 15(1):58–63, 2004.

M. Sonderegger, M. Schumperli, and U. Sauer. Selection of quiescent *Escherichia coli* with high metabolic activity. *Metab. Eng.*, 7:4–9, 2005.

Chapter 7

The yeast metabolome addressed by electrospray ionization mass spectrometry: Initiation of a mass spectral library and its applications for metabolic footprinting by direct infusion mass spectrometry

Jesper Højer-Pedersen, Jørn Smedsgaard, and Jens Nielsen

Metabolomics (in press)

Supplementary material can be found in Appendix B.

Metabolomics
DOI 10.1007/s11306-008-0132-4

ORIGINAL ARTICLE

The yeast metabolome addressed by electrospray ionization mass spectrometry: Initiation of a mass spectral library and its applications for metabolic footprinting by direct infusion mass spectrometry

Jesper Højer-Pedersen · Jørn Smedsgaard · Jens Nielsen

Received: 28 May 2008 / Accepted: 18 September 2008

Abstract Mass spectrometry (MS) has been a major driver for metabolomics, and gas chromatography (GC)-MS has been one of the primary techniques used for microbial metabolomics. The use of liquid chromatography (LC)-MS has however been limited, but electrospray ionization (ESI) is very well suited for ionization of microbial metabolites without any previous derivatization needed. To address the capabilities of ESI-MS in detecting the metabolome of *Saccharomyces cerevisiae*, the *in silico* metabolome of this organism was used as a template to present a theoretical metabolome. This showed that in combination with the specificity of MS up to 84% of the metabolites can be identified in a high mass accuracy ESI-spectrum. A total of 66 metabolites were systematically analyzed by positive and negative ESI-MS/MS with the aim of initiating a spectral library for ESI of microbial metabolites. This systematic analysis gave insight into the ionization and fragmentation characteristics of the different metabolites. With this insight, a small study of metabolic footprinting with ESI-MS demonstrated that biological information can be extracted from footprinting spectra. Statistical analysis of the footprinting data revealed discriminating ions, which could be assigned using the *in silico* metabolome. By this approach metabolic footprinting can advance from a classification method that is used to derive biological information based on guilt-by-association, to a tool for extraction of metabolic differences, which can guide new targeted biological experiments.

Keywords Mass spectrometry · Metabolic footprinting · *Saccharomyces cerevisiae* · Mass spectral library · Metabolic marker

Electronic supplementary material The online version of this article (doi:10.1007/s11306-008-0132-4) contains supplementary material, which is available to authorized users.

J. Højer-Pedersen · J. Smedsgaard · J. Nielsen
Center for Microbial Biotechnology, Technical University of Denmark, Kongens Lyngby, Denmark

J. Nielsen (✉)
Systems Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemivägen 10, 412 96 Gothenburg, Sweden
e-mail: nielsenj@chalmers.se

1 Introduction

Mass spectrometry (MS) has been one of the primary technologies contributing to the continuing advancement of metabolite analysis (Dunn 2008). MS has been refined in the last 20 years and this has enabled today's development of metabolomics, which is aiming at measuring the complete pool of metabolites associated with a cell or an organism. First of all extensive analytical methods are needed, but for metabolomics to succeed in pursuing the goal of global metabolite profiling, efficient data interpretation is required. As many of the analytical strategies for metabolite profiling are based on MS, it is advantageous to compile mass spectra into libraries to facilitate an efficient interpretation of data. The spectral libraries enable efficient identification (dereplication) of compounds, which will speed-up the data digestion and information extraction from metabolite profiles.

There has been a long tradition within GC-MS for construction of mass spectral libraries. While numerous libraries are available with electron impact (EI) spectra (Halket et al. 2005; Schauer et al. 2005) there is a low coverage of spectra for primary metabolites in these libraries. Recently specific emphasis has been put on

construction of biologically relevant libraries that contain metabolite derived spectra, and in this context both sample preparation and derivatization procedures have been addressed. This has placed GC-MS as the preferred analytical techniques for comprehensive and quantitative metabolite profiling for metabolomics (Dunn 2008; Villas-Boas et al. 2005). EI is the primary mode of ionization used for GC-MS and the ionization yields molecular fragments that are specific for the respective compound and represent a fingerprint of the molecule that is well suited for identification (McLafferty and Turecek 1993). Furthermore, the ionization is standardized and very reproducible and this has eased the data collection for libraries that can be applied across instruments and laboratories for decades.

Until now the application of LC-MS for profiling of primary metabolites has been limited. One of the major reasons probably being that no standard reversed phase liquid chromatography method performs very well for a wide range of primary metabolites as these typically are highly polar. Alternatives to standard reversed phase liquid chromatography exist and chromatographic methods like HILIC (hydrophilic interaction chromatography; Nguyen and Schug 2008) or ion-pair chromatography (Coulier et al. 2006) make it possible to separate polar metabolites by LC. The polar properties of primary metabolites make them suited for analysis by electrospray ionization (ESI)-MS. By collecting data for libraries containing ESI spectra, this can ease the interpretation of ESI data from LC-MS or MS fingerprinting. The ESI process is initiated in the liquid phase at atmospheric pressure, and therefore the ESI is affected by the liquid matrix (Matuszewski et al. 2003; Taylor 2005). This gives rise to spectral variation for different sample matrices and makes ESI less robust in terms of spectral reproducibility, because adduct intensities are dependent on e.g., ion strengths (i.e., charge concentration in the liquid phase), ion species, co-elution and design of ion source. This means that sample matrices resulting from sample handling and buffers are reflected in the mass spectrum (Castrillo et al. 2003; Taylor 2005).

ESI will in many cases complement the compound range compared to EI that requires derivatization (van der Werf et al. 2007), and this makes the ESI technique of interest for metabolome analysis as a complementary technique. ESI is a soft ionization technique that only mediates limited molecular fragmentation dependent on instrument operation. Thus, the mass spectrum will primarily contain molecular adduct ions and only provide limited structural information. Fragmentation of ions can be used to gain structural information, and this may be obtained by collision-induced dissociation (CID) either in the ion source region or by tandem MS in a collision cell. Ion source CID is in some cases instrument dependent and resulting mass spectra will also be solvent matrix dependent although standardization can partly reduce this (Bristow et al. 2002). In tandem MS, the ion is pre-selected and fragmented and therefore tandem mass spectra are more specific, and the reproducibility is in general better compared to ion source region fragmentation (Bristow et al. 2004; Gergov et al. 2004; Josephs and Sanders 2004). In addition the sensitivity can be increase by tandem MS since the signal-to-noise ratio (S/N) can be improved dependent on the mass spectrometry technology. The tandem mass spectra can then be used for compound identification by spectral comparison, similarly to identification of EI spectra from GC-MS. The mechanism for fragmentation by CID is caused by collision. The gas phase ions are accelerated and collide with inert gas molecules like nitrogen (N_2) or argon (Ar). The exchange of energy at the time of collision transfers to rotational and translational energy, which causes the fragmentation by bond breakage. The energy required for CID varies dependent on the ion stability, which then make the fragmentation pattern dependent on the fragmentation energy and the partial pressure of the collision gas. The transferability of ESI libraries between instruments is not as good as seen for EI libraries for GC-MS, but Bristow et al. (2004) have successfully created a tandem MS library and evaluated the performance for five different MS instruments.

Applications of ESI tandem MS without a prior separation step (e.g., chromatography or electrophoresis) can be used for direct identification (Boernsen et al. 2005) and even quantification (Nagy et al. 2003) of metabolites in blood samples. Direct infusion ESI-MS has also been used for metabolic footprinting, which is a fingerprinting technique (Fiehn 2002) using the extracellular medium from submerged microbial cultures (Allen et al. 2003; Howell et al. 2006; Kell et al. 2005; Pope et al. 2007). A fingerprinting technique primarily aims at classifying different species and/or samples and can be carried out by MS or various types of spectroscopy e.g., NMR, IR, NIR or fluorescence. The use of ESI-MS provides a fingerprint of the samples, which contains structural information about the metabolites differentiating the samples. The ions in the mass spectrum are directly related to the metabolites in the sample and by identification of ions in the mass spectrum it is possible to extract information about physiological differences rather than just spectral differences assigned by guilt-by-association (Allen et al. 2003). Thus, annotation of the different spectra in ESI-MS will substantially improve the use of this technique for obtaining high-throughput metabolome data that can be used in systems biology.

In an attempt to annotate spectral data from ESI-MS we first approximated the metabolome of *S. cerevisiae* based on the *in silico* metabolic network of this organism (Forster et al. 2003). Hereby we identified the theoretical metabolome that can be measured by MS. We then started to

construct a spectral library containing mass spectra of microbial primary metabolites. This spectral library was initiated with the intension of making it available for the metabolomics community, and hereby allow for a wider use of ESI-MS for metabolome analysis of yeast. The library holds information about the metabolite derived ions observed in positive and negative ESI and the tandem mass spectra resulting from fragmentation of the $[M + H]^+$ and $[M - H]^-$ ions in positive and negative mode, respectively. To validate our approach we performed ESI-MS analysis of the metabolic footprint of different yeast mutants and then used statistics to identify discriminant ions. Some of these ions were then identified using our library and the *in silico* metabolome, and hereby the library was used to directly gain insight into the physiological and metabolic differences of the mutants evaluated.

2 Materials and methods

2.1 *In silico* metabolites

The *in silico* metabolic model of *Saccharomyces cerevisiae*, which was compiled by Forster et al. (2003), was used as a template for generating the theoretical metabolome of *S. cerevisiae*. Furthermore the LIGAND database, which is part of the KEGG database (http://www.kegg.com), was downloaded (March 2006). This contains molecular information about the different metabolites. The intersection between the metabolic model and the LIGAND database was used to obtain a list of defined metabolites for *S. cerevisiae*.

3 Mass spectral library

3.1 Chemicals

Metabolites were all analytical grade and the providers were either Sigma-Aldrich (St. Louis, MO, USA) or Serva (Heidelberg, Germany). Methanol and formic acid used for mass spectrometry were obtained from Sigma-Aldrich (St. Louis, MO, USA). MilliQ-purified water was used for aqueous solutions.

3.2 Metabolite standards

A total of 66 metabolites were prepared in 10 mM aqueous solution and stored at −20°C until analysis. The selected metabolites are listed in Supplementary Material (Table S1) and are primarily from the central carbon metabolism covering a selection of amino acids, organic acids, nucleotides, and glycolytic intermediates.

3.3 MS analysis

All standards were analyzed by nano-ESI-MS using a chip-based Nanomate (Advion BioSciences, Ithaca, NY, USA) robot coupled to a hybrid orthogonal acceleration Q-tof I mass spectrometer with 3.6 GHz TDC detector (Micromass, Manchester, UK).

The Q-tof was tuned to a resolution around 7,000 at full-width half maximum (FWHM) for leucine enkephalin ($C_{28}H_{37}N_5O_7$; $[M + H]^+ = 556.2771$ Da and $[M - H]^- = 554.2615$ Da). In positive mode the mass scale was calibrated using an acidified poly(ethylene glycol) (PEG) solution containing ammonium formate and a fifth-order polynomial mass calibration fit was applied. The calibration in negative mode was obtained by infusion of a PEG bis(carboxymethyl) ether solution, and again a fifth-order polynomial mass calibration fit.

The source block temperature was held at 60°C. High-resolution continuum data was acquired in the mass range of 50–900 Da with a scan time of 1.0 s and interscan of 0.1 s. For positive ESI-MS, the capillary voltage was set to −1.55 kV and 0.35 psi head pressure, and in negative mode the nanospray parameters were set to +1.67 kV and 0.40 psi. Continuum mass spectra were acquired for 1 min at sequentially three different cone voltages of 10 V, 25 V and 40 V. This was done to induce in-source fragmentation and address the ion stability.

Apart from the fragmentation observed by in-source fragmentation caused by elevated cone voltage, the fragmentation patterns were studied by tandem MS. The first quadrupole was set to allow passage of the $[M + H]^+$ and $[M - H]^-$ ions in positive and negative mode, respectively. Nitrogen was used as collision gas in the collision cell at 1.0×10^{-5} mbar and collision energy of 16 eV, 20 eV and 24 eV were applied with cone voltage of 10 V, 25 V and 40 V, respectively.

All standards were diluted to 50 μM with 75% aqueous methanol, and for the analysis in positive ESI-MS the aqueous methanol solution contained 25 mM formic acid as modifier to improve ionization.

The high-resolution continuum data was analyzed by MassLynx 4.0 (Waters, Milford, MA, USA).

4 Metabolic footprinting

4.1 Strain

The *S. cerevisiae* strain BY4741 (*MAT***a** *his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0), BY4741 *hxk2*Δ, and BY4741 *pck1*Δ obtained from the European *S. cerevisiae* archive for functional analysis (EUROSCARF, Frankfurt, Germany) was used in the study (Brachmann et al. 1998). The strains

were auxotrophic and needed histidine, leucine, methionine and uracil, which were available in the medium.

4.2 Medium

The strains were cultivated in a minimal synthetic medium supplemented with a metabolite cocktail according to Allen et al. (2003). The metabolite cocktail contained a selection of amino acids, organic acids and organic bases. The medium was based on yeast nitrogen base (YNB) without amino acids (BD Difco, Franklin Lakes, NJ, USA) and the metabolite cocktail was composed of L-arginine, L-aspartate, L-glutamate, L-histidine, L-leucine, L-lysine, L-methionine, L-serine, L-threonine, L-tryptophan, L-valine, citrate, fumarate, malate, pyruvate, succinate, cytosine and uracil. All compounds from the metabolite cocktail had a concentration of 1 mM in the final medium. Cultivations were performed with glucose, 20 g l^{-1}, as carbon source.

4.3 Pre-culture

Pre-cultures were prepared by inoculating a colony in 5 ml synthetic minimal medium with metabolite cocktail and glucose as carbon source. The pre-cultures were grown for 12 hours in an incubator at 30°C and 150 rpm.

4.4 Shake flask cultivation

Shake flask cultivations were performed in 500-ml baffled flasks with 50 ml medium. The medium was inoculated with the pre-cultures to OD_{600} 0.01 and incubated for 24 h at 30°C and 150 rpm.

4.5 Mass spectrometry analysis

The samples taken from the shake flasks were filtered immediately after sampling to obtain the metabolic footprint. Prior to analysis the samples were diluted 10 times by methanol. Samples of 1 µl were injected via a FAMOS autosampler (LC packings, The Netherlands) into a carrier flow of methanol at 15 µl min^{-1} from a microflow binary pump (Agilent Technologies, Santa Clara, CA, USA). Before reaching the mass spectrometer additionally 5 µl min^{-1} of 2.0% (v/v) aqueous formic acid was added through a T-piece using a syringe pump. The total flow of 20 µl min^{-1} was injected into a Micromass Q-tof I mass spectrometer with 3.6 GHz TDC detector (Manchester, UK) equipped with a Z-spray ion sources.

The Q-tof was tuned in positive mode as described above for the mass spectral library. For analysis, the desolvation and nebulizer gas flow was 320 l h^{-1} and 20 l h^{-1}, respectively. The source block temperature was set to 80°C and the desolvation temperature was 100°C. The capillary voltage was 3,200 V and the cone voltage was 17 V. Scans were collected from m/z 40 to 500 with a scan time of 2.0 s and 0.1 s interscans delay.

4.6 Data analysis

The mass spectra were converted by in-house written software and normalized by the total sum of ion intensities. The spectra were aligned in bins of 0.1-Da width between 40 and 500 Da. The data was organized into at matrix with each row representing a sample and the columns containing the variables corresponding to the binned masses from the mass spectrum. All variables (columns in the data matrix) only containing zeros were removed from the data. Principal component analysis (PCA) was used to visualize the data (Jolliffe 2002). Before PCA the individual variables were centered and scaled to unit variance. Analysis of variance (ANOVA) was used to select group specific variables. The data analysis was executed in MATLAB (MathWorks, Natick, MA, USA).

5 Results and discussion

5.1 *In silico* yeast metabolites

The metabolic network of the yeast *S. cerevisiae* has been extensively studied in the last decade using various flux models of different complexity. Based on the genome sequence, Forster et al. (2003) compiled a genome-scale metabolic model of *S. cerevisiae*, which contained approximately 1,200 metabolic reactions involving around 550 metabolites. The metabolites contained in the model were used to identify the theoretical metabolome of *S. cerevisiae*. In order to obtain chemical structures and molecular masses the LIGAND part of the KEGG database (http://www.kegg.com), which contains 13,785 metabolite entries (March 2006), was used. By extracting the intersection of the around 550 *in silico* metabolites from the metabolic model and of the compound list from KEGG, it was possible to compile a list of 438 unique metabolites with known (and exact) molecular formula, from which the accurate (monoisotopic) mass could be calculated (Supplementary Material, Table S2). Here it should be noted that the genome scale model of *S. cerevisiae* does not cover detailed metabolic pathways for lipids, and basically no lipids are therefore included in the metabolite list.

Of the 438 metabolites 366 (84%) molecular formulas are unique in terms of elemental composition and ideally possible to distinguish by MS simply based on the accurate mass. Deriving the nominal mass of these unique formulas

gives rise to 268 (61%) distinct nominal masses. And therefore high resolution mass spectrometers will be superior; since they can distinguish many more metabolites compared to low resolution instruments with unit mass resolution. Although some metabolites may have the exact same elemental composition and thereby mass, in some cases tandem MS can be used to distinguish these due to differences in fragmentation of these molecules. For complex mixtures analyzed by direction infusion tandem MS, a mixture of ions may enter the collision cell and result in mixed tandem MS spectra, since the selection of precursor ions in the first quadrupole is at nominal mass. The mixed tandem MS spectra will composite of a mixture of fragmented ions dependent on the abundance on the individual precursor ions. Hence, the ability to distinguish metabolites with the same nominal mass will require separation before tandem MS detection.

The mass distribution of metabolites is illustrated in the histogram in Fig. 1. Here it is observed that there is a high density of metabolites having a molecular mass between 100 and 300 Da in accordance with the metabolite mass distribution presented by Kell et al. (2005). Further the atomic composition for the metabolites was analyzed and reported in Table 1. Carbon, hydrogen and oxygen are present in almost all metabolites. Nitrogen is present in about 2/3 of the metabolites and phosphor is found in 39%. Sulphur is contained in 10% of the metabolites. Especially, the isotope pattern of sulphur might be helpful for identification of this part of the metabolome with MS. Iron is rarely found, but the isotope pattern will be characteristic for these metabolites.

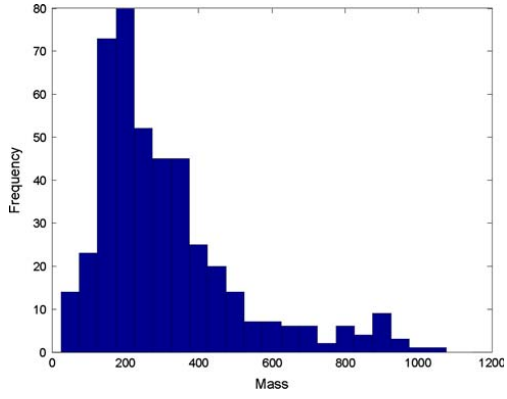

Fig. 1 Histogram showing the distribution of molecular masses for metabolites present in the *in silico* metabolome of yeast

Table 1 Frequency of atoms in metabolites

Atom	In # of metabolites	Percentage
C	430	98
H	436	99
N	283	65
O	426	97
S	46	11
P	172	39
Fe	3	1

6 ESI-MS/MS library

In order to establish an initial library, 66 metabolites from the core of the central metabolism were selected. Of the metabolites in the library, 27 were associated with amino acid metabolism, 11 were related to the TCA and the glyoxylate cycle, 5 related to glycolysis, 4 involved in fatty acid metabolism, 1 from the pentose phosphate pathway, 2 related to redox metabolism, 4 related to galactose and starch metabolism, and 5 were nucleotides.

The diluted metabolite standards were analyzed one by one. The sensitivity of detection of the individual metabolites varied dependent on ionization polarity and cone voltage. The analysis of the metabolite standards revealed that positive and negative electrospray complement each other very well. They favor different metabolites, but in combination they can ease the identification. There are no rules without exceptions, but as expected, amino acids and generally metabolites containing amino groups are detected better by positive ESI, whereas organic acids and phosphorylated compounds were commonly better detected by negative ESI when quantifying the $[M + H]^+$ and $[M - H]^-$ detection.

6.1 Positive mode

It was observed that the compounds—apart from formation of different adduct ions (Table 2)—can ion-exchange acidic protons with positive metal ions and especially Na^+. This was observed for metabolites having one or several carboxylic acid groups, but also phosphorylated metabolites showed extensive ion-exchange. In some cases two or three protons were exchanged by sodium. Therefore, the observation of ion-exchange patterns is helpful in terms of ion identification and can indicate the number of acidic protons in the compound.

In this experiment the cone voltage was set to sequentially be 10 V, 25 V and 40 V during MS analysis of the individual metabolites. The optimal cone voltage was primarily found to correlate with the molecular mass, so that the higher the mass the higher the cone voltage. Up to m/z

Table 2 Frequently observed ions for the metabolites in the spectral library

	Positive ESI		Negative ESI	
	Ion composition	Mass change	Ion composition	Mass change
Adducts	$[M + H]^+$	1	$[M + Cl]^-$	+35[a]
	$[M + Na]^+$	23		
	$[M - H + 2Na]^+$	45		
	$[M - 2H + 3Na]^+$	67		
Fragments	$[M - NH_3 + H]^+$	−16	$[M - H]^-$	−1
	$[M - H_2O + H]^+$	−17	$[M - NH_3 - H]^-$	−18
	$[M - CHONH_2 + H]^+$	−44	$[M - H_2O - H]^-$	−19
	$[M - CHOOH + H]^+$	−45	$[M - CO_2 - H]^-$	−45
			$[M - H_3PO_4 - H]^-$	−99
Multimers	$[2\ M + H]^+$	+M + 1	$[2\ M - H]^-$	+M − 1
	$[2\ M + Na]^+$	+M + 23		

[a] Adduct ions have characteristic isotope pattern according to the adduct molecule

around 300 Da/e, a cone voltage of 25 V was found to give the best detection, and above m/z 300 Da/e 40 V was preferred. None of the metabolites were well detected at 10 V.

Amino acids were best detected in positive mode, probably due to the amino group. The typical base peak ion was $[M + H]^+$. During tandem MS loss of formic acid (CHOOH) was observed and when an additional amino group, apart from the α-amino group, was present in the molecule, then one amino group was typically lost as ammonia (Rogalewicz et al. 2000).

An extensive formation of sodium adducts appeared for carboxylic acids. In Fig. 2a the positive mass spectrum for citramalic acid is presented. Both the $[M + H]^+$ and the $[M + Na]^+$ ion were found in the spectrum and here the sodium adduct was the larger of those, but the ratio between the ions varies dependent of the sodium concentration in the solution used for ESI-MS analysis. Further inspection of the spectrum in Fig. 2a shows a dimer sodium adduct ($[2\ M + Na]^+$) at mass 319.10 Da/e. Apart from molecular adduct ions one clear fragment from in-source CID at mass 103.05 Da/e was seen. This ion corresponds to the loss of formic acid from the protonated molecular ion. The fragmentation was confirmed by tandem MS, where loss of formic acid gave the predominant ion (Table S3). Additionally, loss of water and a combined loss of water and formic acid was observed by tandem MS. The loss of water (H_2O) and formic acid (CHOOH) were the main neutral losses found by tandem MS of carboxylic acids.

For nucleotides containing one or multiple phosphates several ion adducts were observed (Fig. 3). The multiple adducts arose from ion exchange of acidic protons with primarily sodium. From the chemical structure, GTP (guanosine triphosphate) seemingly contains four acidic protons associated with the triphosphate; but from the mass spectrum in Fig. 3 an ion corresponding to $[M - 5H + 6Na]^+$ was present. The fifth proton might have been recruited from the guanine part of GTP, where a resonance form of the purine derived rings might have bounded a sodium ion instead of a proton. The mass spectrum

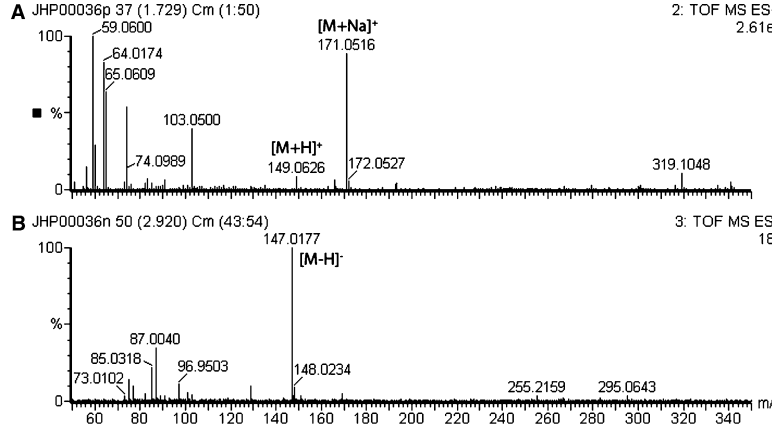

Fig. 2 ESI-MS of citramalic acid (M = 148.0372 Da). In positive mode (**a**) both the protonated and sodiated molecular ion is observed. Negative ESI-MS (**b**) is found to mainly produce the deprotonated molecular ion

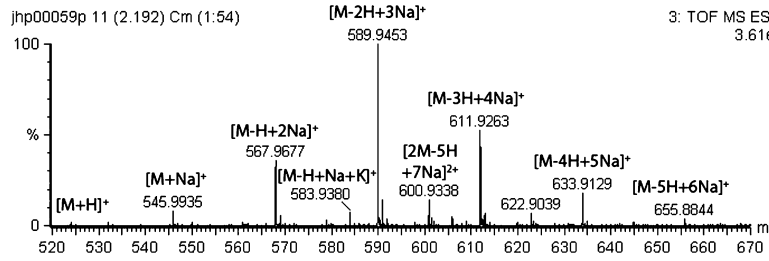

Fig. 3 Positive ESI-MS of GTP (guanosine triphosphate). The major ions are assigned and shows extensive adduct formation with sodium. A combined sodium and potassium adduct is also present

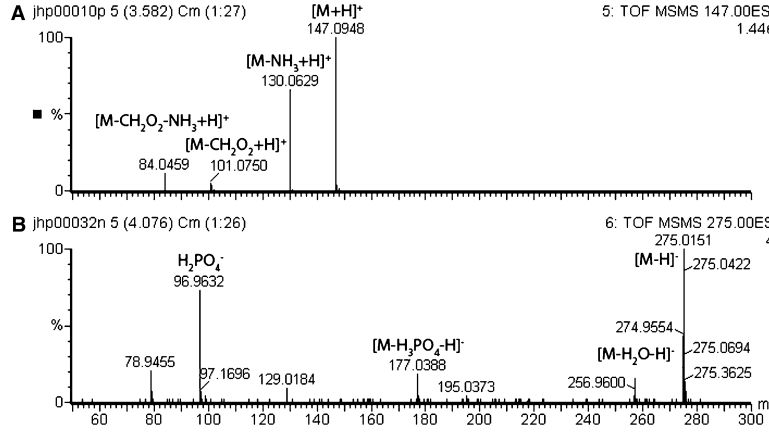

Fig. 4 Tandem mass spectrometry. **a** Fragmentation of the $[M + H]^+$ ion of glutamine (M = 146.0691 Da) giving three fragments which are assigned according to their loss of mass. **b** Fragmentation of the $[M - H]^-$ ion of 6-phosphogluconic acid (M = 276.0246 Da). Note that the $H_2PO_4^-$ ion is a characteristic fragment in negative mode for phosphate-containing compounds

for ATP (data not shown) did not contain an $[M - 5H + 6Na]^+$ ion, but up to $[M - 4H + 5Na]^+$ matching the four acidic protons from the triphosphate. The high degree of sodium-adduct formation in both cases probably originated from the metabolite standard solutions, which were based on the sodium salt of the compounds. In biological samples the sodium concentration would presumably be lower and therefore the adducts with high exchange of sodium would be less frequent, but trace amount of sodium will almost always cause sodium adduct formation in positive ESI.

The cone voltage determined the acceleration of the ions in the ion-source and thereby the fragmentation by in-source CID. The in-source CID should be minimized for tandem MS, since the CID preferably should occur in the collision cell after "isolation" of the precursor ion. In the current experimental setup for MS analysis, the cone voltage was confounded with the collision energy. This means that for low cone voltage (10 V) low collision energy (16 eV) was used and vice versa. This was based on the assumption that the more stable the ion, the higher the cone voltage and thus the higher the collision energy. Standard conditions for tandem MS could not be derived from the data. The fragmentation energy needed to fragment the precursor ion was metabolite dependent. Rather a setup with ramping fragmentation energy is suggested to obtain the tandem mass spectrum. An example of fragmentation is shown in Fig. 4a, where $[M + H]^+$ of glutamine is fragmented at 20 eV. A very clean mass spectrum with practically no noise was obtained. The fragments corresponded to loss of ammonia and formic acid plus a combination of both. All fragmentation resulted from positive mode are summarized in Supplementary Material, Table S3. The CID revealed no gas phase reactions with radical ions in comparison with fragmentation from EI ionization. This made the spectra simpler, but also less specific. However, specificity was gained by accurate mass fragments.

6.2 Negative mode

It was difficult to get a stable nanospray in negative mode. Loss or drop-out of signal/spray happened in many cases and caused a reduced data quality for the negative ESI results. Therefore, the library is sparse and not optimal in negative mode.

In contrast to positive mode, no correlation between molecular mass and cone voltage was found. Rather the choice of cone voltage seemed to be compound specific.

E.g., citric acid was very labile in negative ESI and the low cone voltage of 10 V was optimal for the $[M - H]^-$ ion. At higher cone voltage the de-protonated molecular ion exhibited pronounced fragmentation. The lack of correlation between molecular mass and cone voltage in negative mode might be caused by the ionization mechanism, which requires breakage of covalent bonds to produce $[M - H]^-$ ions. The energy needed for bond breakage is compound specific and not just related to the mass, e.g., it requires less energy to ionize a carboxylic acid than a carbohydrate in negative mode.

Compared to positive ESI, the mass spectra were generally simpler in negative ESI, since the molecules mainly produced $[M - H]^-$ ions. This makes negative ESI suited for determination of molecular mass preferably in combination with positive ESI. Figure 2b shows the negative ESI spectrum of citramalic acid and a clear peak corresponding to the de-protonated molecular ion is seen at 147.02 Da/e. Tandem MS confirmed that the ion of 87.00 Da/e was a fragment of citramalic acid, which corresponds to the loss of acetic acid. Figure 4b shows an example of tandem MS of 6-phosphogluconic acid fragmented at 24 eV. First of all the fragment of 96.96 Da/e was a strong indication of a phosphate containing compound. Secondly, the fragment of 177.04 Da/e corresponded to the loss of phosphoric acid from the precursor ion. Except for these phosphate-related fragments there was only limited structural information available in the tandem MS spectrum. Loss of water from the precursor ion was observed. A list of all CID results in negative mode is summarized in Supplementary Material, Table S4.

7 Metabolite identification

Combination of the knowledge of ionization (Table 2) and the metabolite list corresponding to the theoretical metabolome (Table S2) serves as a good starting point for identification of unknown metabolites in mass spectra. Before starting to search the metabolite list, it is important to consider which adducts and fragments can be expected. To illustrate this, take e.g., a mass spectrum in positive mode, in which an unknown ion of mass 203.06 Da/e is found. First of all it relevant to confirm that it is a single charged ion by looking at the neighboring isotope ions in the mass spectrum. If the ions are separated by approximately 1.0 Da, the ion is single-charged. Secondly, a number of adduct ions are common in positive mode (Table 2). If the ion was an $[M + H]^+$ ion the M would have mass 202.05 and similarly 185.03 for $[M + NH_4]^+$, 180.07 for $[M + Na]^+$, 164.10 for $[M + K]^+$, and 158.09 for $[M - H + 2Na]^+$. Assuming a mass precision of ± 0.1 Da, it is possible to suggest the metabolites for ion 203.06 in Table 3 using Table S2 from Supplementary Material. This reduces the possibilities to 9 different elemental compositions, because fructose, galactose, glucose, inositol, mannose and sorbose; and fucose and rhamnose have the same elemental composition, respectively. The final identification requires additional information. The suggested metabolite/molecule can be further supported by adding *a priori* knowledge. If a suggested metabolite contains sulfur, then the isotope pattern for sulfur should be reflected in the mass spectrum. Similarly, if the ion is a potassium adduct the isotope pattern can reveal it.

Table 3 Tentative metabolites for ion 203.06 ± 0.1 Da

Ion	M	Tentative identification
$[M + H]^+$	202.05 ± 0.1	None
$[M + NH4]^+$	185.03 ± 0.1	3-phosphoserine
		2-amino-3-carboxymuconate semialdehyde
$[M + Na]^+$	180.07 ± 0.1	Fructose
		Galactose
		Glucose
		Inositol
		Mannose
		Sorbose
		Hydroxyphenylpyruvate
$[M + K]^+$	164.10 ± 0.1	Phenylpyruvate Fucose
		Rhamnose
		S-methyl-L-methionine
$[M - H + 2Na]^+$	158.09 ± 0.1	Dihydrooroate allantoin
		2-isopropylmaleate

8 An example of metabolic footprinting

The principal of footprinting for *S. cerevisiae* was presented by Allen et al. (2003) and is closely related to fingerprinting especially in terms of the analytical approach, in which high-throughput techniques are used. Fingerprints are directed towards intracellular metabolites and footprints aim at covering the extracellular metabolites. A fingerprint is a snapshot of a metabolic state in contrast to footprints that are resulting from integrated changes over time as a function of substrate uptake and secretion.

The footprints contain integrated information which eventually might strengthen a characteristic response and reveal a biomarker. The distinction between finger- and footprinting is not stringent (in the literature) and probably the term fingerprinting is the most frequently used to cover the approach. For this study the sampling of extracellular metabolites was used, since this is much simpler compared

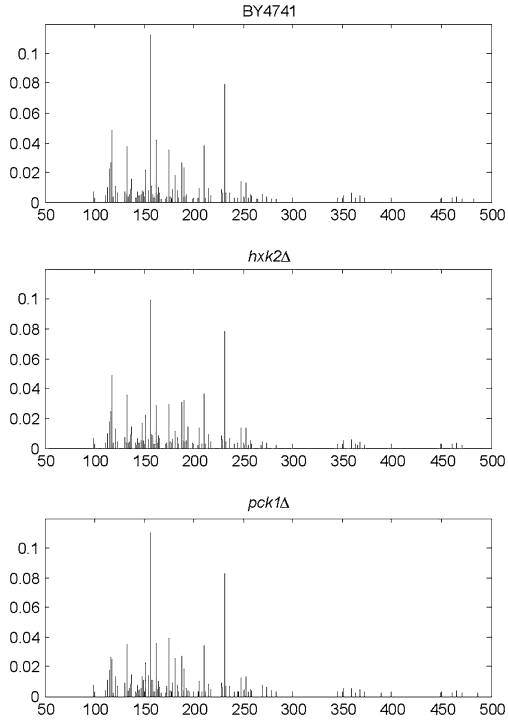

Fig. 5 Examples of metabolic footprints obtained by direct infusion mass spectrometry. The mass spectra represent the BY4741, *hxk2*Δ and *pck1*Δ strain, respectively

to sampling for intracellular metabolites, and because the extracellular metabolites are easily accessible and allow high-throughput (Kell et al. 2005).

In the present example we employed direct infusion ESI-MS for metabolite footprinting. The MS analysis is used to study the metabolite profiles of a reference strain and two single gene deletion mutants of *S. cerevisiae* carrying deletions in genes encoding enzymes in the central carbon metabolism.

The shake flask cultures were grown for 24 h and samples were taken for MS analysis. Each strain were cultured in triplicate (biological replicates) and three samples were taken from each culture (analytical replicates)—giving a total of 27 data points. Typical mass spectra from the three different strains are shown in Fig. 5. The spectra are very similar and not readily possible to distinguish by the naked eye.

The data was aligned in order to make the data accessible for multivariate analysis. Alignment was obtained by overlaying a grid with 0.1 mass units to bin the data, which yielded 159 non-zero variables (masses). By subjecting the

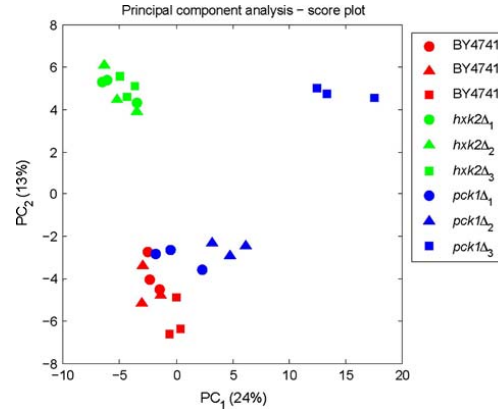

Fig. 6 Principal component analysis (PCA) of mass spectrometry footprinting data. The score plot shows the first two principal components (PCs) describing 37% of the data variation

processed data to PCA, the score plot in Fig. 6 showed that strain specific information was available in the data. The first and second principal component (PC) explained 24% and 13% of the data variation, respectively. There was high similarity for the samples from the *hxk2*Δ strain, which clearly clustered in the loading plot and the variability between samples could not be distinguished from the analytical replicates. Similar observations were found for the reference strain BY4741. The same data consistency was not observed for the *pck1*Δ strain although the analytical replicates are alike. Especially one of the biological replicates of *pck1*Δ seemed to be rather different from the two others. This might indicate the presence of an outlier, but it was difficult to justify an omission of the sample ($pck1\Delta_3$) when the remaining group was defined by only 6 data points from 2 biological replicates with 159 variables. Therefore, the data analysis was done both with and without the most extreme sample ($pck1\Delta_3$) of the *pck1*Δ mutant. Repeating the PCA without $pck1\Delta_3$ gave the score plot in Fig. 7. This PCA was capable of capturing 86% of the overall variation by the two-first components, thus representing the full data variation much better. Still the data for the *hxk2*Δ strain formed a cluster and this mutant was separated from the two others by the first PC. The second PC explained 30% of the data variation and partly accounted for the differences between BY4741 and *pck1*Δ, but there seemed to be a striking similarity between the *pck1*Δ1 sample and all the BY4741 samples.

Analysis of variance (ANOVA) was used to address the differences between the *S. cerevisiae* strains. A histogram of the *P*-values from the ANOVA is presented in Fig. 8 and reveals that there was a considerable number of variables for which $P < 0.05$. To inspect the low *P*-values the

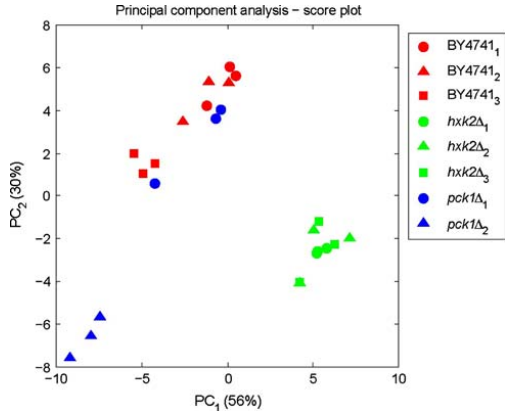

Fig. 7 PCA of MS footprinting data excluding the *pck1*Δ_3 sample. The score plot shows the first two principal components (PCs) describing 86% of the data variation

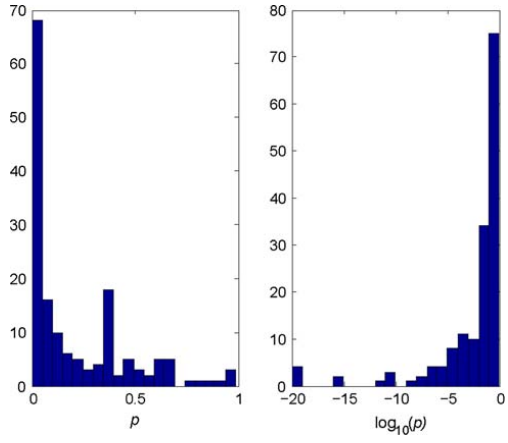

Fig. 8 Distribution of *P*-values found from analysis of variance (ANOVA) of footprinting data

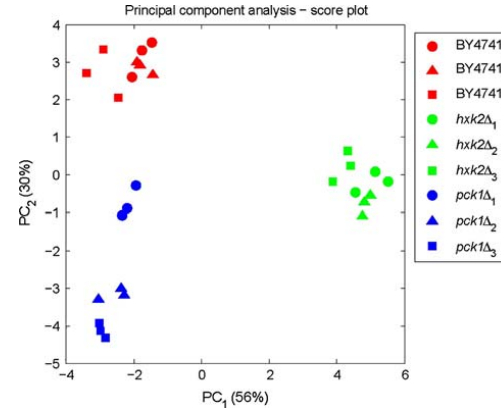

Fig. 9 PCA of footprinting data after selection of significant variables by analysis of variance (ANOVA)

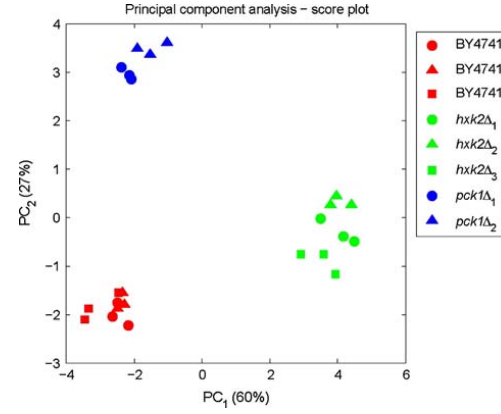

Fig. 10 PCA of footprinting data excluding the *pck1*Δ_3 sample after selection of significant variables by ANOVA

histogram to the right in Fig. 8 shows the distribution of $\log_{10}(P)$, and from this histogram it is seen that some of the variables were highly significant. The most significant variables were selected for investigation, since these might represent physiological markers for the individual mutants. A very conservative level was set and only variables with $P < 10^{-5}$ were chosen, which reduced the number of variables from 159 to 21.

Applying PCA to the data reduced by ANOVA confirmed that the remaining variables were highly group-specific. Figures 9 and 10 present the score plots for the data after ANOVA with and without the *pck1*Δ_3 sample, respectively. These score plots were accounting for 86–87% of the total variance and confirmed that the variables left after ANOVA were highly strain specific, since data clustered according to the strain. However, in Fig. 9 with all samples including the suspected outlier *pck1*Δ_3 there were still some divergence among the *pck1*Δ samples. The cluster is not as delimited as for the two other strains.

Strain specific variables are listed in Table 4, which presents the most significant ions found from the ANOVA. To gain more specific information regarding the strain differences, tentative identification of ions was done based on the ion mass using Table S2, and this resulted in the list of metabolites that represent the different strains given in Table 4. The errors listed in Table 4 indicate relatively poor mass accuracy for a tof instrument. The poor accuracy

Table 4 Highly significant ions ($P < 10^{-5}$)

Mass	Metabolite	Adduct	Error (Da)
131.06	Itaconic acid	$[M + H]^+$	0.03
132.09	Leucine/isoleucine	$[M + H]^+$	−0.01
150.04	Methionine	$[M + H]^+$	−0.02
177.06	Isopropylmalate	$[M + H]^+$	−0.02
194.08	Isopropylmalate	$[M + NH4]^+$	−0.02
148.03	Glutamate	$[M + H]^+$	−0.03
159.06	Isopropylmaleate	$[M + H]^+$	−0.01
190.03	Pyridoxal	$[M + Na]^+$	−0.02
115.03	Glycerol	$[M + Na]^+$	−0.01
118.08	Valine	$[M + H]^+$	−0.01
206.03	Pyridoxal	$[M + K]^+$	0.01
158.08	Threonine	$[M + K]^+$	0.06
162.08	Nicotinate	$[M + K]^+$	0.08
119.08	Succinate	$[M + H]^+$	0.05
179.1	Cys–Gly	$[M + H]^+$	0.05
110.07	Glycerol	$[M + NH4]^+$	−0.01
230.98	Citrate	$[M + K]^+$	−0.01
199.01	Isopropylmalate	$[M + Na]^+$	−0.05
147.06	Glutamine	$[M + H]^+$	−0.02
190.09	*N*-Acetyl-glutamate	$[M + H]^+$	0.02
273.88	?		

The ions are assigned with a tentative metabolite suggested based on the mass ± 0.1 Da (Table S2)

might be caused by all ions in Table 4 have a relatively low mass compared to the calibration range by PEG, thus the calibration might not provide optimal accuracy below 200 Da/e. To obtain a more confident identification the samples should ideally have been reanalyzed by tandem MS of the significant ions.

The results from the footprinting experiment confirmed that strain characteristics can be derived from metabolic footprints (Allen et al. 2003; Kell et al. 2005). This indicated the strength of metabolites as functional markers, where the genome sets the boundaries for metabolic pathways, which will be inherited downstream to the metabolites and then be detected in the metabolic footprints. Even the initial PCA revealed the presence of major clusters in the underlying MS data and the clusters were strain associated (Figs. 6 and 7). The *hxk2*Δ strain segregated from the two other strains, which was also expected based on the physiological knowledge of the strain. *HXK2* encodes hexokinase II in *S. cerevisiae* and plays a role in carbon catabolite repression (Entian and Mecke 1982). The *hxk2*Δ mutant has a distinct physiology compared to the reference strain. It grows at a slower rate on glucose compared to the reference strain and no fermentative products are formed in the initial exponential phase, which leads to higher biomass yield. The diauxic shift is also shorter, which indicates altered carbon catabolite repression, since the enzymes for respirative growth are not repressed (Diderich et al. 2001). Although slower growth of the *hxk2*Δ strain, all strains were sampled after exhaustion of glucose, which was the primary carbon source. Thus, the considerable differences detected for the *hxk2*Δ strain was not caused by residual glucose.

Variable selection was required to clearly distinguish the reference and the *pck1*Δ strain by PCA. ANOVA successfully selected variables (ions), which accounted for these differences, and led to a complete separation of the three strains (Figs. 9 and 10). It was expected that the *pck1*Δ strain would be alike the reference strain. *PCK1* is not known to have any regulatory role on any other genes, but is solely acting as an enzyme. *PCK1* encodes for phosphoenolpyruvate carboxykinase catalyzing the reaction of oxaloacetate to phosphoenolpyruvate. The reaction is part of the gluconeogenesis and primarily active during growth on 2-C carbon sources like ethanol and acetate. The *pck1*Δ mutant is not viable on ethanol (McCammon 1996). As this experiment was carried out on glucose, a high similarity between BY4741 and the *pck1*Δ mutant was expected, which was also observed. However, minor differences seemed to be present, which shows the specificity and sensitivity of metabolic footprinting for phenotypic characterization.

9 Metabolic markers

Metabolic footprinting was proposed as a guilt-by-association approach for functional genomics (Allen et al. 2003); but to take footprinting beyond the point of guilt-by-association, information regarding exact metabolic differences must be extracted from the footprints. In this experiment, the highly significant ions that were selected by ANOVA are interesting metabolic markers, which might explain the genetic and metabolic differences. Many of the ions listed in Table 4 were contained in the footprinting medium; however, the uptake of these was different for the different strains, which resulted in significantly different quantities in the footprints.

The tentative metabolites were mapped on to the metabolic pathways of *S. cerevisiae* to see if there was convergence in parts of metabolism. According to KEGG (http://www.kegg.com) leucine/isoleucine, isopropylmalate, isopropylmaleate and valine are all associated with the biosynthesis/degradation of valine, leucine and isoleucine. Valine and leucine are pyruvate-derived amino acids (Stryer 1995), which could point to metabolic differences around the pyruvate branch point. Thus, the footprints were found to contain metabolic characteristics for the compared strains and the results supported the specificity of

footprinting. Even deletion of a presumably inactive enzyme (*PCK1*) during growth on glucose gave distinct footprints.

10 Concluding remarks

The building of a spectral library for ESI-MS of microbial metabolites has been initiated. Although the mass spectrum library only covers a minor part of the metabolites, still it represents a core set of metabolites and it can obviously be expanded in the future. The library contains information about ion formation in electrospray for a selection of primary metabolites. The library also contains fundamental information regarding fragmentation patterns for different classes of metabolites, which makes it possible to predict candidate ions for other and similar metabolites. Together with the theoretical metabolome, it has provided insight to the biological information contained in metabolic footprints. With this approach metabolic footprinting by ESI-MS can reach beyond the principle of guilt-by-association and metabolic markers can be extracted from the high-throughput data. The findings can be used to setup hypotheses for further biological experiments that are targeted and more specific.

Acknowledgements The authors wish to thank Gerald Hofmann for fruitful discussions and comments to this manuscript.

References

Allen, J., Davey, H. M., Broadhurst, D., Heald, J. K., Rowland, J. J., Oliver, S. G., et al. (2003). High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnology, 21*, 692–696. doi:10.1038/nbt823.

Boernsen, K. O., Gatzek, S., & Imbert, G. (2005). Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Analytical Chemistry, 77*, 7255–7264. doi:10.1021/ac0508604.

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., et al. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England), 14*, 115–132. doi:10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

Bristow, A. W. T., Nichols, W. F., Webb, K. S., & Conway, B. (2002). Evaluation of protocols for reproducible electrospray in-source collisionally induced dissociation on various liquid chromatography/mass spectrometry instruments and the development of spectral libraries. *Rapid Communications in Mass Spectrometry, 16*, 2374–2386. doi:10.1002/rcm.843.

Bristow, A. W. T., Webb, K. S., Lubben, A. T., & Halket, J. (2004). Reproducible product-ion tandem mass spectra on various liquid chromatography/mass spectrometry instruments for the development of spectral libraries. *Rapid Communications in Mass Spectrometry, 18*, 1447–1454. doi:10.1002/rcm.1492.

Castrillo, J. I., Hayes, A., Mohammed, S., Gaskell, S. J., & Oliver, S. G. (2003). An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry, 62*, 929–937. doi:10.1016/S0031-9422(02)00713-6.

Coulier, L., Bas, R., Jespersen, S., Verheij, E. R., van der Werf, M. J., & Hankemeier, T. (2006). Simultaneous quantitative analysis of metabolites using ion pair-liquid chromatography–electrospray ionization mass spectrometry. *Analytical Chemistry, 78*, 6573–6582. doi:10.1021/ac0607616.

Diderich, J. A., Raamsdonk, L. M., Kruckeberg, A. L., Berden, J. A., & van Dam, K. (2001). Physiological properties of *Saccharomyces cerevisiae* from which hexokinase II has been deleted. *Applied and Environmental Microbiology, 67*, 1587–1593. doi:10.1128/AEM.67.4.1587-1593.2001.

Dunn, W. B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Physical Biology, 5*, 1–24. doi:10.1088/1478-3975/5/1/011001.

Entian, K.-D., & Mecke, D. (1982). Genetic evidence for a role of hexokinase isozyme PII in carbon catabolite repression in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry, 257*, 870–874.

Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Molecular Biology, 48*, 155–171. doi:10.1023/A:1013713905833.

Forster, J., Famili, I., Fu, P., Palsson, B. O., & Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research, 13*, 244–253. doi:10.1101/gr.234503.

Gergov, M., Weinmann, W., Meriluoto, J., Uusitalo, J., & Ojanpera, I. (2004). Comparison of product ion spectra obtained by liquid chromatography/triple-quadrupole mass spectrometry for library search. *Rapid Communications in Mass Spectrometry, 18*, 1039–1046. doi:10.1002/rcm.1445.

Halket, J. M., Waterman, D., Przyborowska, A. M., Patel, R. K., Fraser, P. D., & Bramley, P. M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of Experimental Botany, 56*, 219–243. doi:10.1093/jxb/eri069.

Howell, K. S., Cozzolino, D., Bartowsky, E. J., Fleet, G. H., & Henschke, P. A. (2006). Metabolic profiling as a tool for revealing Saccharomyces interactions during wine fermentation. *FEMS Yeast Research, 6*, 91–101. doi:10.1111/j.1567-1364.2005.00010.x.

Jolliffe, I. T. (2002). *Principal component analysis*. New York: Springer Verlag.

Josephs, J. L., & Sanders, M. (2004). Creation and comparison of MS/MS spectral libraries using quadrupole ion trap and triple-quadrupole mass spectrometers. *Rapid Communications in Mass Spectrometry, 18*, 743–759. doi:10.1002/rcm.1402.

Kell, D. B., Brown, M., Davey, H. M., Dunn, W. B., Spasic, I., & Oliver, S. G. (2005). Metabolic footprinting and systems biology: The medium is the message. *Nature Reviews. Microbiology, 3*, 557–565. doi:10.1038/nrmicro1177.

Matuszewski, B. K., Constanzer, M. L., & Chavez-Eng, C. M. (2003). Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Analytical Chemistry, 75*, 3019–3030. doi:10.1021/ac020361s.

McCammon, M. T. (1996). Mutants of *Saccharomyces cerevisiae* with defects in acetate metabolism: Isolation and characterization of Acn- mutants. *Genetics, 144*, 57–69.

McLafferty, F. W., & Turecek, F. (1993). *Interpretation of mass spectra*. Sausalito: University Science Books.

Nagy, K., Takats, Z., Pollreisz, F., Szabo, T., & Vekey, K. (2003). Direct tandem mass spectrometric analysis of amino acids in dried blood spots without chemical derivatization for neonatal screening. *Rapid Communications in Mass Spectrometry, 17*, 983–990. doi:10.1002/rcm.1000.

Nguyen, H. P., & Schug, K. A. (2008). The advantages of ESI-MS detection in conjunction with HILIC mode separations: Fundamentals and applications. *Journal of Separation Science, 31*, 1465–1480. doi:10.1002/jssc.200700630.

Pope, G. A., MacKenzie, D. A., Defemez, M., Aroso, M. A. M. M., Fuller, L. J., Mellon, F. A., et al. (2007). Metabolic footprinting as a tool for discriminating between brewing yeasts. *Yeast (Chichester, England), 24*, 667–679. doi:10.1002/yea.1499.

Rogalewicz, F., Hoppilliard, Y., & Ohanessian, G. (2000). Fragmentation mechanisms of α-amino acids protonated under electrospray ionization: A collisional activation and ab initio theoretical study. *International Journal of Mass Spectrometry, 195*, 565–590. doi:10.1016/S1387-3806(99)00225-0.

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., et al. (2005). GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Letters, 579*, 1332–1337. doi:10.1016/j.febslet.2005.01.029.

Stryer, L. (1995). *Biochemistry*. New York: W. H Freemand and Company.

Taylor, P. J. (2005). Matrix effects: The Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clinical Biochemistry, 38*, 328–334. doi:10.1016/j.clinbiochem.2004.11.007.

Villas-Boas, S. G., Mas, S., Akesson, M., Smedsgaard, J., & Nielsen, J. (2005). Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews, 24*, 613–646. doi:10.1002/mas.20032.

van der Werf, M. J., Overkamp, K. M., Muilwijk, B., Coulier, L., & Hankemeier, T. (2007). Microbial metabolomics: Toward a platform with full metabolome coverage. *Analytical Biochemistry, 370*, 17–25. doi:10.1016/j.ab.2007.07.022.

Chapter 8

Discussion and perspectives

The metabolomics research is still at its infancy compared to transcriptomics and proteomics, but continuously new techniques and tools are published and progress is evident (Dunn, 2008; Mashego et al., 2007). There are a continuously more articles published and the annual number of publication from 1998 and onwards still seems to increase.

To conclude the thesis, this final chapter addresses Chapters 5–7 in the context of the field of metabolomics and some of the most recent publications. Sample preparation is pivotal for metabolomics and the most recent results from literature will be discussed in the context of Chapter 5. Furthermore, design of experiments will be addressed with the ability to conduct multi-parallel experiments. Then the impact of mass spectrometry (MS) on metabolomics is discussed, and finally the chapter is ended by a discussion of the dynamic responses and the concept of dynamic footprinting.

8.1 Sample preparation

Since the publication of Chapter 5 (Publication 1) in 2005 other research groups have investigated sampling and extraction of microorganisms, and the paper has been cited numerous times. Bolten et al. (2007) studied the sampling of several Gram-positive and Gram-negative bacteria and found that quenching in cold aqueous methanol lead to unspecific leakage and significant losses of metabolites. Fast filtration was tested as another option and found suited for analysis of metabolites with low turnover rates (*e.g.* amino acids and TCA cycle intermediates). Further Winder et al. (2008) investigated the quenching and extraction of metabolites from *Escherichia coli* with an approach comparable to the study

for *Saccharomyces cerevisiae* in Chapter 5. Winder et al. (2008) suggested using cold aqueous methanol for quenching although leakage occurred and advocated to monitor the exo-metabolome and supernatants from the quenching step. Loret et al. (2007) also evaluated quenching of *S. cerevisiae* in cold aqueous methanol and estimated the leakage to be in the range of 5%, when the contact time with methanol was minimized. The results from Bolten et al. (2007), Loret et al. (2007) and Winder et al. (2008) support the findings in Chapter 5 that quenching in cold aqueous methanol causes leakage. Bacteria seem more sensitive to methanol quenching compared to yeasts, but it was shown in Chapter 5 that the cell wall of *S. cerevisiae* was damaged. An alternative to cold aqueous methanol quenching could be quenching in cold saline-glycerol, which was proposed by Villas-Boas and Bruheim (2007) for both bacteria and yeasts. Glycerol is a cryoprotectant and often used for strain preservation, thus it is a very reasonable choice of quenching solution as it might not harm the cells.

Extraction with methanol using freeze-thaw cycles seemed to be the preferred extraction method for *E. coli* (Winder et al., 2008) as was found for *S. cerevisiae* (Chapter 5) for metabolomics. However, Loret et al. (2007) re-validated the extraction method for *S. cerevisiae* using boiling ethanol; and the paper must be partly considered as a comment to Chapter 5 as they strongly disagree with the indications that boiling ethanol might not be optimal for metabolome extraction. The paper (Loret et al., 2007) is rather a defence of extraction by boiling ethanol than comparison of methods, but Loret et al. (2007) shows that boiling ethanol is a very valid extraction method for phosphorylated metabolites and some organic acids. There are only limited loses of these metabolites during boiling in three minutes in the absence of cells. Furthermore, it is correctly stated by Loret et al. (2007) that the poor recoveries of sugar phosphates found in Chapter 5 could be caused by the choice of analytical method (GC-MS), which was already addressed in the Chapter 5. However, chloroform-methanol-buffer extraction was found to be the best method for extraction of sugar phosphates in Chapter 5, hence a comparison between boiling ethanol and chloroform-methanol-buffer extraction would have been appreciated. Another reason for discrepancies could be the protocol for extraction, since Loret et al. (2007) showed that it was important to evaporate the extract with the cells, which was not done in Chapter 5.

It is important to keep in mind that the two studies (Loret et al. (2007) and Chapter 5) do not have the same scope, since Loret et al. (2007) are pursuing a targeted approach whereas Chapter 5 is aiming at global metabolite profiling (metabolomics). Taking the extraction results for *E. coli* from Winder et al. (2008) methanol with freeze-thaw cycles and boiling ethanol may work equally well, and Winder et al. (2008) also conclude that complementary extraction methods will be needed to cover the whole metabolome.

In continuation of extraction, I think it is worthwhile to mention that the ultimate goal of metabolomics is to achieve quantification of the full metabolome, and quantitative rather than qualitative information about the presence of metabolites is essential for microbial metabolomics to succeed. There are several examples of quantification of intracellular metabolites concentration (*e.g.* Buchholz et al., 2002; Lange et al., 2001; Theobald et al., 1993), but it is questionable whether it is meaningful to talk about an intracellular concentration. The cytosol is not a well-mixed liquid with diffusion being the primary mass transport for metabolites in the cytosol. Of course some metabolites present in high amounts may be almost equally distributed in the cell due to evenly distribution of involved metabolic enzymes in the cytosol. The kinetics of the glycolytic (EMP) pathway has successfully between modeled based on measured intracellular concentrations, but this pathway must be considered as the highway in the metabolism. For other pathways channeling is needed for some reactions to proceed based on thermodynamic calculations (Stephanopoulos et al., 1998), *e.g.* the galactose pathway (de Jongh et al., submitted), and this underlines that the cytosol cannot be considered as well-mixed. Rather metabolite quantities should be reported relative to the cell mass.

Metabolome data is highly affected by the actual experimental setup. Not only the growth conditions are influencing your results, but sampling and sample preparation are critical parameters for your results (Chapter 5). The choice of methods will influence the final data sets, and data sets from different experiments and labs are likely to be confounded, which at present makes inter-laboratorial comparison of data difficult. Initiatives for microarrays (Brazma et al., 2001) and protein analysis (Orchard et al., 2003) have already been presented – what is the minimum experimental information needed to go along with the data to make it usable for reanalysis by other research groups? Likewise Lindon et al. (2005) have recommended guidelines for standardization and reporting of metabolic analyzes, which partly initiated the Metabolomics Standard Initiative. The essence is that genome sequences are more or less conserved independently of the experimental setup. However, this is not the case for downstream components such as RNAs, proteins and metabolites. These are all context-specific and reflect individual experimental conditions and physiological states. For microbes the Metabolomics Standard Initiative has materialized into a specific set of requirements for reporting of microbial metabolomics (van der Werf et al., 2007b).

8.2 Experimental design

Metabolites carry integrated information from the surroundings (ecotype) and the genes (genotype) as discussed by Smedsgaard and Nielsen (2005), and the results from Chapter 6 show that changing the primary carbon source results in a significant change of the exo-metabolome. This clearly illustrates the impact of the ecotype. The dimension of ecotype for *S. cerevisiae* has, however, not been studied extensively using metabolomics; particularly in comparison with functional studies of genotypes, *e.g.* in a study by Giaever et al. (2002) who analyzed many single gene deletion mutants growing on different media. Manipulation of the ecotype makes it possible to activate different parts of metabolism, which might reveal pleiotropy of genes. This should definitely have more focus in metabolomics.

The results in Chapter 6 with cultivation in microtiter plates (MTPs) also show that throughput can be significantly increased by using MTPs compared to ordinary shake flasks. High-throughput is a prerequisite for screening of several strains and not only can the genotype be an experimental factor but the ecotype should be varied by using different growth media as already discussed. Furthermore, MTPs opens up for the possibility of conducting several repetitions of the same experiment, which is essential to obtain valid results that are statistically significant. The experimental design is critical for the ability to produce valid data in metabolomics (Dunn, 2008). The experimental design should cover not only the biological experiment but also the following analysis to be able to estimate the uncertainty of measurements and avoid bias. Problems arising from improper experimental design are discussed in Section 8.4.

8.3 Mass spectrometry and metabolomics

The central position of MS in metabolomics is clear (Dunn, 2008; Dettmer et al., 2007; Villas-Boas et al., 2005) and I am confident that the constant improvement of MS instruments will keep MS in that position for future metabolite studies. The high-resolution instruments becoming available today can determine the mass within less than 1 ppm (Breitling et al., 2006). These instruments allow high mass resolving power (Sleno et al., 2005), which combined with database searches (Kind and Fiehn, 2006) will continue to drive the research of metabolomics forward. This should be seen in combination with the continuing improvement of sensitivity and the ability to cover a large range of dynamic response. Besides improved MS instruments, there is a need for efficient separation and whether this separation will be obtained by gas or liquid chromatography (van der Werf et al., 2007a) or capillary electrophoresis (Montona and Soga, 2007) is hard to say — maybe a combination of these is needed.

Enormous amounts of information are available from MS data and at present I see that the deconvolution and analysis of data determine the pace for metabolomics. Ute Rossner (personal communication) stated that as a rule-of-thump, one should expect to use 3 times the data acquisition time to analyze the data. To speed up the data analysis reference data for metabolites are important. As already discussed in Chapter 7, there is a long tradition for collection of spectral libraries for GC-MS, whereas the availability of the spectral libraries for electrospray ionization (ESI)-MS for microbial metabolites is limited. Hence, it is very valuable to start collecting tandem MS data. Not only does an ESI tandem MS library contain specific information about the individual metabolites, but it can also be used to identify characteristic fragmentation patterns, which might be used for identification of unknown metabolites.

For data analysis theoretical metabolomes are available through genome-scale metabolic models for various organisms and these represent a set (or subset) of metabolites that potentially can be detected by metabolome analysis. van der Werf et al. (2007a) took the union of three microbial *in silico* metabolomes (*E. coli*, *Bacillus subtilis*, and *S. cerevisiae*) of which 399 metabolites could be purchased and used for the evaluation of a platform for metabolome analysis. This platform enabled analysis of 380 out of 399 metabolites using a combination of GC-MS and LC-MS methods. In Chapter 7, the *in silico* metabolome of *S. cerevisiae* was used to investigate the capabilities of detecting it by ESI-MS. This theoretical analysis emphasized why MS has had and still has such a central position in metabolomics: MS has the specificity required to undertake the challenge of metabolomics.

8.4 Dynamic footprinting

To conclude, this final Section is used to discuss perspectives of investigating correlations in metabolomics. Correlation between intracellular metabolite pools has been demonstrated by Arkin et al. (1997), and correlations have even been observed in cases where the metabolites are far from being directly linked by metabolic pathways (Steuer et al., 2003). The intracellular metabolite concentration change rapidly and react to environmental changes (de Koning and van Dam, 1992), and therefore the dynamic behavior of the system have small time constants, which will require very frequent sampling to appropriately describe the dynamics of system. *E.g.* the response of the concentration of intermediates in the glycolysis on a glucose pulse has been extensively studied in *S. cerevisiae*. Here, samples are taken within few seconds to make it possible to access the kinetics of the pathway (Lange et al., 2001; Theobald et al., 1993, 1997; Visser et al., 2002).

In the research areas of omics, the focus has until now primarily been on

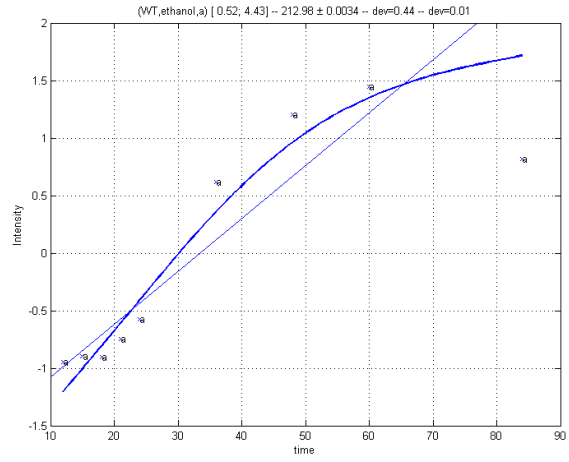

Figure 8.1 Observations from dynamic footprinting *i.e.* the evolution of metabolic footprints. From metabolic footprints obtained by ESI-MS the intensity of some ions was found to change over time as the yeast grew. The present example initially shows an exponential increase in intensity corresponding to excretion of a metabolite, which is expected from a shaken batch cultivation. At a later stage in the end of the batch cultivation the intensity tends to decrease, which indicates that the yeast has started to consume the metabolite after a metabolic shift.

measurement of concentrations or presence of components, *e.g.* genes, mRNA and proteins, to distinguish differences at different steady states. However, life is mainly driven by fluxes (Nielsen, 2003; Sauer, 2004). Hence, in my opinion the analysis of data from time series will give insight into the interplay between the different functional levels in the cell. This will make it possible to reveal cascade reactions, which are hard to derive from steady-state observations. In this context dynamic footprinting may be one way to address the dynamics of the metabolome.

The footprinting approach (Kell et al., 2005) has proved it strengths (*e.g.* Chapters 6 and 7); but might be sensitive to differences in growth rates and time of sampling. Allen et al. (2003) showed that the footprints changed over time and from a principal component analysis (PCA) their trajectory were reproducible. At the same time Allen et al. (2003) argued that no major changes of the footprint would occur when a culture had reached the stationary phase. However, this may be questionable for *e.g. S. cerevisiae* grown on glucose, since glucose is mainly converted to ethanol, which subsequently can be consumed after the diauxic shift. Although no significant growth might be observed after depletion

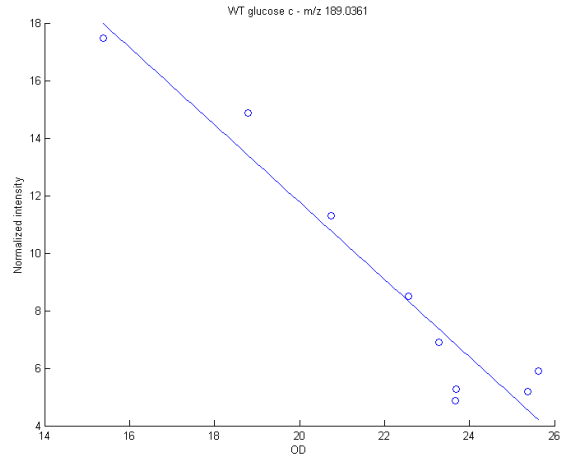

Figure 8.2 Growth-correlated trends from dynamic footprinting. Along with the sampling for metabolic footprinting the biomass concentration was measured by optical density (OD). The OD and ion intensity are correlated for some ions, which demonstrates that the changes observed in the dynamic footprints are related to the growth of yeast. The correlation is interpreted to correspond to constant yield coefficients on biomass, which represents a physiological feature that can be extracted from the dynamic footprints.

of glucose, ethanol can slowly be consumed and changes will occur in the culture broth (footprint). Secondly, if the cells are kept without nutrients for a while, they might undergo lysis and intracellular metabolites will contribute to the footprint.

To address the changes in metabolic footprints, we propose dynamic metabolic footprinting by analyzing time series of samples of extracellular metabolites to identify physiologic characteristics that reaches beyond differences in specific growth rate. At present the setup for dynamic footprinting is not fully tuned, which explains why only limited data and details are presented, but preliminary results indicate the presence of information in the dynamic footprints. We know from Allen et al. (2003) that footprints change over time, so at first changes over time were sought to pinpoint the dynamic ions (Figure 8.1). Dependent on the medium composition a number of metabolites will be taken up and consumed from the medium, which would result in a declining concentration over time. Also some metabolites will be excreted over time as seen in Figure 8.1, where the data points form an S-curve that tends to level off at the end. Here, the metabolite is first produced and later on consumed during the course of a batch cultivation.

During growth several metabolites will be consumed and excreted as a result

of growth, and the changes are very likely to correlate with growth. Figure 8.2 shows an example of such trends between ion intensity and optical density that is used to measure the biomass concentration. The plot shows a linear trend with negative correlation corresponding to uptake of the metabolite. Both positive and negative correlations between extracellular metabolites and biomass are observed from batch cultures. These correlations represent constant yield coefficients during the batch culture. The yield coefficients (Y) are defined by the ratio between two rates (r) of the i'th and j'th component (Nielsen et al., 2002):

$$Y_{ij} = \frac{r_j}{r_i} \tag{8.1}$$

which for constant yield coefficients also can be derived from concentrations (c):

$$Y_{ij} = \frac{\Delta c_j}{\Delta c_i} \tag{8.2}$$

Similar to the growth correlated ions, pairwise comparison of the individual ion intensities in Figure 8.3 shows that some of them are linearly correlated. The plot is an example of consistent observations from three independent biological replicates.

The data has been acquired by direct infusion MS, which might not be considered as a quantitative approach by default, but the present results demonstrated at least semi-quantitative responses in the MS data. The ion intensities were found to vary as the yeast was growing (Figure 8.2) and showed profiles over time which could be expected for shaken batch cultures (Figure 8.1). To maintain the quantitative aspect it is of course important that the sample matrix remains similar throughout the sampling period, since direct infusion MS is known to be sensitive to matrix effect.

To develop the concept of dynamic footprinting further there are currently some challenges that need to be dealt with — or better practices that should be considered. Design of experiments and randomization is of outmost importance to enable valid conclusions from this kind of data. Of course this applies for all experiments, but partly lack of these considerations was a problem for interpreting a large data set that I received from post doc Sandrine Mas, who had worked in the group of Jens Nielsen. Secondly, it is important that the sampling is frequent enough to establish the dynamic response and represent it with sufficient data points. Another issue is the number of replicates needed in the data set to extract biological information. There has been a tendency to use three replicates for biological experiments, but it is questionable whether this is sufficient to adequately detect outliers. The current data set contains triplicates of shake flask cultures, and thereby it suffers from having too few replicates to

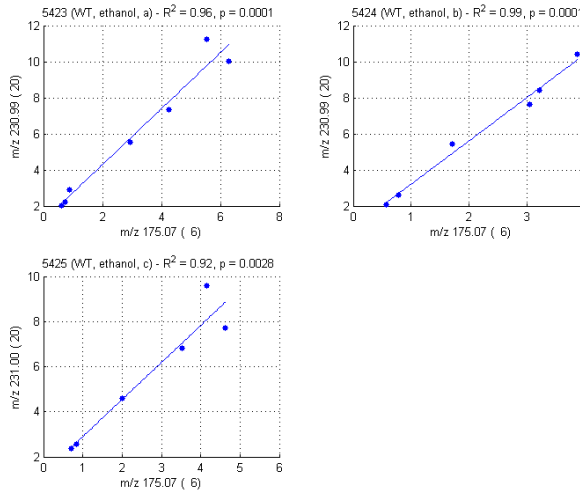

Figure 8.3 Correlation between ions. Pairwise correlations between intensities of ions in the dynamic footprint were identified and these represent yield coefficients between two metabolites.

identify outliers and obtaining confident statistics, which in general weakens the data analysis. Furthermore, there are no analytical replicates. Unfortunately, the sample list was not randomized before analysis, which caused sample groups to be confounded with instrument drift over time. All in all this has lead to a heavily biased data set with promising indication without allowing proof of concept.

Instrument drift and bias from the MS analysis will be very likely, when many samples have to be analyzed, and normalization is needed to correct for these effects. The normalization is not trivial and both normalization within and between spectra is needed. Part of the normalization could be accomplished by using an internal standard, but in the present data set, variation in signal-to-noise ratios had an impact on the number of ions detected with low intensities, which resulted in suboptimal performance of the subsequent data analysis.

In conclusion, this data demonstrates that dynamic footprinting will be exciting to pursue further. It is a non-targeted approach for phenotypic feature extraction, which links to fluxes and flux ratios. Hence, the approach is capable of depicting physiological changes, which is the key objective in many biological studies. Dynamic footprinting should be considered as a screening tool that opens up new possibilities for high-throughput screening.

References

J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.*, 21(6):692–696, 2003.

A. Arkin, P. D. Shen, and J. Ross. A test case of correlation metric construction of a reaction pathway from measurements. *Science*, 277(5330):1275–1279, 1997.

C. J. Bolten, P. Kiefer, F. Letisse, J.-C. Portais, and C. Wittmann. Sampling for metabolome analysis of microorganisms. *Anal. Chem.*, 79(10):3843–3849, 2007.

A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat. Genet.*, 29(4):365–371, 2001.

R. Breitling, A. R. Pitt, and M. P. Barrett. Precision mapping of the metabolome. *Trends in Biotechnol.*, 24(12):543–548, 2006.

A. Buchholz, J. Hurlebaus, C. Wandrey, and R. Takors. Metabolomics: quantification of intracellular metabolite dynamics. *Biomol. Eng.*, 19(1):5–15, 2002.

W. de Koning and K. van Dam. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal. Biochem.*, 204(1):118–123, 1992.

K. Dettmer, P. A. Aronov, and B. D. Hammock. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.*, 26:51–78, 2007.

W. B. Dunn. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.*, 5(1):011001, 2008.

G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. E. Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-MacDonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis,

and M. Johnston. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418(6896):387–391, 2002.

D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic, and S. G. Oliver. Metabolic footprinting and systems biology: The medium is the message. *Nat. Rev. Microbiol.*, 3(7):557–565, 2005.

T. Kind and O. Fiehn. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*, 7:234, 2006.

H. C. Lange, M. Eman, G. van Zuijlen, D. Visser, J. C. van Dam, J. Frank, M. J. de Mattos, and J. J. Heijnen. Improved rapid sampling for *in vivo* kinetics of intracellular metabolites in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.*, 75(4):406–415, 2001.

J. C. Lindon, J. K. Nicholson, E. Holmes, H. C. Keun, A. Craig, J. T. M. Pearce, S. J. Bruce, N. Hardy, S. A. Sansone, H. Antti, P. Jonsson, C. Daykin, M. Navarange, R. D. Beger, E. R. Verheij, A. Amberg, D. Baunsgaard, G. H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J. N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I. D. Wilson, M. D. Reily, D. G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer, and M. Spraul. Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol.*, 23(7):833–838, 2005.

M. O. Loret, L. Pedersen, and J. Francois. Revised procedures for yeast metabolites extraction: application to a glucose pulse to carbon-limited yeast cultures, which reveals a transient activation of the purine salvage pathway. *Yeast*, 24(1):47–60, 2007.

M. R. Mashego, K. Rumbold, M. D. Mey, E. Vandamme, W. Soetaert, and J. J. Heijnen. Microbial metabolomics: past, present and future methodologies. *Biotechnol. Lett.*, 29(1):1–16, 2007.

M. R. N. Montona and T. Soga. Metabolome analysis by capillary electrophoresis–mass spectrometry. *J. Chromatogr. A*, 1168(1–2):237–246, 2007.

J. Nielsen. It is all about metabolic fluxes. *J. Bacteriol.*, 185(24):7031–7035, 2003.

J. Nielsen, J. Villadsen, and G. Liden. *Bioreaction engineering principles*. Kluwer cademic/Plenum Publishers, New York, 2nd edition, 2002.

S. Orchard, H. Hermjakob, and R. Apweiler. The proteomics standards initiative. *Proteomics*, 3(7):1374–1376, 2003.

U. Sauer. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.*, 15(1):58–63, 2004.

L. Sleno, D. A. Volmer, and A. G. Marshall. Assigning product ions from complex MS/MS spectra: The importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.*, 16(2):183–198, 2005.

J. Smedsgaard and J. Nielsen. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J. Exp. Bot.*, 56(410):273–286, 2005.

G. N. Stephanopoulos, A. A. Aristidou, and J. Nielsen. *Metabolic engineering. Principles and methodologies.* Academic Press, San Diego, 1st edition, 1998.

R. Steuer, J. Kurths, O. Fiehn, and W. Weckwerth. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*, 19(8):1019–1026, 2003.

U. Theobald, W. Mailinger, M. Reuss, and M. Rizzi. *In vivo* analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique. *Anal. Biochem.*, 214(1):31–37, 1993.

U. Theobald, W. Mailinger, M. Baltes, M. Rizzi, and M. Reuss. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnol. Bioeng.*, 55(2):305–316, 1997.

M. J. van der Werf, K. M. Overkamp, B. Muilwijk, L. Coulier, and T. Hankemeier. Microbial metabolomics: toward a platform with full metabolome coverage. *Anal. Biochem.*, 370(1):17–25, 2007a.

M. J. van der Werf, R. Takors, J. Smedsgaard, J. Nielsen, T. Ferenci, J. C. Portais, C. Wittmann, M. Hooks, A. Tomassini, M. Oldiges, J. Fostel, and U. Sauer. Standard reporting requirements for biological samples in metabolomics experiments: microbial and in vitro biology experiments. *Metabolomics*, 3:189–194, 2007b.

S. G. Villas-Boas and P. Bruheim. Cold glycerol–saline: The promising quenching solution for accurate intracellular metabolite analysis of microbial cells. *Anal. Biochem.*, 370(1):87–97, 2007.

S. G. Villas-Boas, S. Mas, M. Akesson, J. Smedsgaard, and J. Nielsen. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.*, 24(5):613–646, 2005.

D. Visser, G. A. van Zuylen, J. C. van Dam, A. Oudshoorn, M. R. Eman, C. Ras, W. M. van Gulik, J. Frank, G. W. van Dedem, and J. J. Heijnen. Rapid sampling for analysis of *in vivo* kinetics using the BioScope: a system for continuous-pulse experiments. *Biotechnol. Bioeng.*, 79(6):674–681, 2002.

C. L. Winder, W. B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G. M. Stephens, and R. Goodacre. Global metabolic profiling of *Escherichia coli* cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. *Anal. Chem.*, 80(8):2939–2948, 2008.

Appendices

Appendix A

Atomic isotopes

The monoisotopic mass of a molecule is calculated based the atomic mass of the individual isotopes. Table A.1 lists the mass of isotopes that typically are present in biological molecules. Additionally, the abundances of isotopes are presented for the calculation of isotope patterns.

References

J. R. de Laeter, J. K. Bohlke, P. D. Bievre, H. Hidaka, H. S. Peiser, K. J. R. Rosman, and P. D. P. Taylor. Atomic weights of the elements: Review 2000 - (IUPAC technical report). *Pure and Applied Chemistry*, 75(6):683–800, 2003.

Table A.1 Stable mass isotopes and their natural abundances

Atom	Isotope	Mass (Da)	Abundance (%)	Relative abundance
Hydrogen	^{1}H	1.007825	99.9885	100.0
	^{2}H	2.014102	0.0115	0.0
Lithium	^{6}Li	6.015122	7.59	8.2
	^{7}Li	7.016004	92.41	100.0
Carbon[a]	^{12}C	12.000000	98.93	100.0
	^{13}C	13.003355	1.07	1.1
Nitrogen	^{14}N	14.003074	99.636	100.0
	^{15}N	15.000109	0.364	0.4
Oxygen	^{16}O	15.994915	99.757	100.0
	^{17}O	16.999132	0.038	0.0
	^{18}O	17.999160	0.205	0.2
Sodium	^{23}Na	22.989770	100.00	100.0
Phosphorus	^{31}P	30.973761	100.00	100.0
Sulphur	^{32}S	31.972071	94.99	100.0
	^{33}S	32.971458	0.75	0.8
	^{34}S	33.967867	4.25	4.5
	^{36}S	35.967081	0.01	0.0
Chloride	^{35}Cl	34.968853	75.76	100.0
	^{37}Cl	36.965903	24.24	32.0
Potassium	^{39}K	38.963707	93.2581	100.0
	^{40}K	39.963999	0.0117	0.0
	^{41}K	40.961826	6.7302	7.2
Bromide	^{79}Br	78.918338	50.7	100.0
	^{81}Br	80.916291	49.3	97.2

[a] The mass of the ^{12}C isotope is exactly 12 Da by definition

Reference: de Laeter et al. (2003)

Appendix B

Supplementary material for Chapter 7

Table S1 contains an index of metabolites analyzed for the mass spectral library.

Table S2 contains the metabolome of yeast based on the *in silico* model of *Saccharomyces cerevisiae*.

Table S3 summarizes fragmentation by positive tandem MS of metabolites included in the mass spectral library.

Table S4 summarizes fragmentation by negative tandem MS of metabolites included in the mass spectral library.

ESM Table S1

Click here to download Table: S1_Library_index.xls

@BCL@E00B8BB3.xls

#	Compound name	KEGG compound ID	Formula	Monoisotopic mass
00001	4-Aminobutyric acid	C00334	C4H9NO2	103.063
00002	Glycine	C00037	C2H5NO2	75.032
00003	L-Alanine	C00041	C3H7NO2	89.048
00004	L-Arginine	C00062	C6H14N4O2	174.112
00005	L-Aspartic acid	C00049	C4H7NO4	133.038
00006	L-Asparagine	C00152	C4H8N2O3	132.054
00008	L-Cysteine	C00097	C3H7NO2S	121.020
00009	L-Glutamic acid	C00025	C5H9NO4	147.053
00010	L-Glutamine	C00064	C5H10N2O3	146.069
00011	L-Histidine	C00135	C6H9N3O2	155.070
00012	L-Homocystine	C01817	C8H16N2O4S2	268.055
00014	L-Isoleucine	C00407	C6H13NO2	131.095
00015	L-Leucine	C00123	C6H13NO2	131.095
00016	L-Lysine	C00047	C6H14N2O2	146.106
00017	L-Methionine	C00073	C5H11NO2S	149.051
00019	L-Phenylalanine	C00079	C9H11NO2	165.079
00020	L-Proline	C00148	C5H9NO2	115.063
00021	L-Serine	C00065	C3H7NO3	105.043
00023	L-Tryptophan	C00078	C11H12N2O2	204.090
00024	L-Tyrosine	C00082	C9H11NO3	181.074
00025	L-Valine	C00183	C5H11NO2	117.079
00027	O-acetyl-L-serine	C00979	C5H9NO4	147.053
00028	O-Phospho-L-serine	C01005	C3H8NO6P	185.009
00029	N-Acetyl-L-glutamic acid	C00624	C7H11NO5	189.064
00030	Betaine	C00719	C5H11NO2	117.079
00031	2-Hydroxyisobutyric acid	-	C4H8O3	104.047
00032	6-phosphogluconic acid	C00345	C6H13O10P	276.025
00033	Apidic acid	C06104	C6H10O4	146.058
00034	cis-Aconitic acid	C00417	C6H6O6	174.016
00035	Citraconic acid	C02226	C5H6O4	130.027
00036	Citramalic acid	C00815	C5H8O5	148.037
00037	Citric acid	C00158	C6H8O7	192.027
00038	DL-Isocitric acid	C00311	C6H8O7	192.027
00039	Fumaric acid	C00122	C4H4O4	116.011
00040	Glutaric acid	C00489	C5H8O4	132.042
00041	Glyceric acid	C00258	C3H6O4	106.027
00042	Glyoxylic acid	C00048	C2H2O3	74.000
00043	Itaconic acid	C00490	C5H6O4	130.027
00044	Lactic acid	C00186	C3H6O3	90.032
00045	Malic acid	C00149	C4H6O5	134.022
00046	Myristic acid	C06424	C14H28O2	228.209
00047	Oxalacetic acid	C00036	C4H4O5	132.006
00048	Oxalic acid	C00209	C2H2O4	89.995
00049	Pimelic acid	C02656	C7H12O4	160.074
00050	Tartaric acid	C00898	C4H6O6	150.016
00051	α-Ketoapidic acid	C00846	C6H8O5	160.037
00052	Flavin adenine dinucleotide (FAD)	C00016	C27H33N9O15	785.157
00054	NAD	C00003	C21H28N7O14	664.117
00055	ATP	C00002	C10H16N5O13	506.996
00057	AMP	C00020	C10H14N5O7P	347.063
00059	GTP	C00044	C10H16N5O14	522.991

@BCL@E00B8BB3.xls

#	Compound name	KEGG compound ID	Formula	Monoisotopic mass
00061	cGMP	C00942	C10H12N5O7P	345.047
00063	Cytosine	C00380	C4H5N3O	111.043
00064	Coenzyme A	C00010	C21H36N7O16	767.115
00065	Acetyl-DL-carnitine	C02571	C9H18NO4	204.124
00066	Fructose 1,6-bisphosphate	C00354	C6H14O12P2	339.996
00067	Phospho(enol)pyruvate	C00074	C3H5O6P	167.982
00068	3-phosphoglyceric acid	C00197	C3H7O7P	185.993
00069	2-phosphoglyceric acid	C00631	C3H7O7P	185.993
00071	Glucose 1-phosphate	C00103	C6H13O9P	260.030
00072	Galactose 6-phosphate	C01113	C6H13O9P	260.030
00074	Uridine 5'-diphospho-glucose	C00029	C15H24N2O17	566.055
00075	Uridine 5'-diphospho-galactose	C00052	C15H24N2O17	566.055
00076	Trehalose	C01083	C12H22O11	342.116
00079	N-acetyl-D-glucosamine	C00140	C8H15NO6	221.090
00081	Riboflavin	C00255	C17H20N4O6	376.138

ESM Table S2

Click here to download Table: S2_Yeast_metabolite_list.xls

@BCL@E00B0CC3.xls

#	KEGG ID	Metabolite	Formula	Monoisotopic mass
1	C00084	Acetaldehyde	C2H4O	44.026215
2	C00033	Acetate	C2H4O2	60.021129
3	C00164	Acetoacetate	C4H6O3	102.031694
4	C00332	Acetoacetyl-CoA	C25H40N7O18P3S	851.136338
5	C00900	2-Acetolactate	C5H8O4	132.042259
6	C00024	Acetyl-CoA	C23H38N7O17P3S	809.125773
7	C00147	Adenine	C5H5N5	135.054495
8	C00212	Adenosine	C10H13N5O4	267.096754
9	C00224	Adenylylsulfate	C10H14N5O10PS	427.019899
10	C00008	ADP	C10H15N5O10P2	427.029415
11	C00301	ADPribose	C15H24N5O14P2	560.079499
12	C00099	beta-Alanine	C3H7NO2	89.047678
13	C00041	L-Alanine	C3H7NO2	89.047678
14	C00499	Allantoate	C4H8N4O4	176.054555
15	C01551	Allantoin	C4H6N4O3	158.04399
16	C04556	4-Amino-2-methyl-5-phosphomethylpyrimidine	C6H10N3O4P	219.040892
17	C04409	2-Amino-3-carboxymuconate semialdehyde	C7H7NO5	185.032422
18	C11355	4-amino-4-deoxychorismate	C10H11NO5	225.063722
19	C04874	2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8-dihydropteridine	C9H13N5O4	255.096754
20	C01300	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	C7H9N5O2	195.075625
21	C01279	4-Amino-5-hydroxymethyl-2-methylpyrimidine	C6H9N3O	139.074562
22	C04454	5-Amino-6-(5'-phosphoribitylamino)uracil	C9H17N4O9P	356.073315
23	C01268	5-Amino-6-(5'-phosphoribosylamino)uracil	C9H15N4O9P	354.057665
24	C01268	5-Amino-6-(ribosylamino)-2,4-(1H,3H)-pyrimidinedione 5'-phosphate	C9H15N4O9P	354.057665
25	C04807	2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine	C7H11N5O8P2	355.008285
26	C01092	8-Amino-7-oxononanoate	C9H17NO3	187.120843
27	C00956	L-2-Aminoadipate	C6H11NO4	161.068808
28	C04076	L-2-Aminoadipate 6-semialdehyde	C6H11NO3	145.073893
29	C00568	4-Aminobenzoate	C7H7NO2	137.047678
30	C00334	4-Aminobutanoate	C4H9NO2	103.063329
31	C00555	4-Aminobutyraldehyde	C4H9NO	87.068414
32	C05715	gamma-Amino-gamma-cyanobutanoate	C5H8N2O2	128.058578
33	C03373	Aminoimidazole ribotide	C8H14N3O7P	295.056936
34	C00430	5-Aminolevulinate	C5H9NO3	131.058243
35	C02220	2-Aminomuconate	C6H7NO4	157.037508
36	C03824	2-Aminomuconate 6-semialdehyde	C6H7NO3	141.042593
37	C02229	3-Aminopropanal	C3H7NO	73.052764
38	C05714	alpha-Aminopropiononitrile	C3H6N2	70.053098
39	C00020	AMP	C10H14N5O7P	347.063084
40	C00108	Anthranilate	C7H7NO2	137.047678
41	C00652	D-Arabinono-1,4-lactone	C5H8O5	148.037173
42	C00216	D-Arabinose	C5H10O5	150.052823
43	C00062	L-Arginine	C6H14N4O2	174.111676
44	C00152	L-Asparagine	C4H8N2O3	132.053492
45	C00049	L-Aspartate	C4H7NO4	133.037508
46	C00441	L-Aspartate 4-semialdehyde	C4H7NO3	117.042593
47	C00002	ATP	C10H16N5O13P3	506.995745
48	C00120	Biotin	C10H16N2O3S	244.088163
49	C01159	2,3-Bisphospho-D-glycerate	C3H8O10P2	265.959269
50	C04002	But-1-ene-1,2,4-tricarboxylate	C7H8O6	188.032088
51	C00169	Carbamoyl phosphate	CH4NO5P	140.982709
52	C04692	2-[3-Carboxy-3-(methylammonio)propyl]-L-histidine	C11H19N4O4	271.14063
53	C04441	2-(3-Carboxy-3-aminopropyl)-L-histidine	C10H16N4O4	256.117155
54	C04236	3-Carboxy-4-methyl-2-oxopentanoate	C7H10O5	174.052823
55	C01063	6-Carboxyhexanoyl-CoA	C28H46N7O19P3S	909.178203
56	C01302	1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate	C12H16NO9P	349.056268
57	C01269	5-O-(1-Carboxyvinyl)-3-phosphoshikimate	C10H13O10P	324.024633
58	C00487	Carnitine	C7H16NO3	162.113018
59	C00112	CDP	C9H15N3O11P2	403.018181
60	C00307	CDPcholine	C14H27N4O11P2	489.115156
61	C00570	CDPethanolamine	C11H20N4O11P2	446.060381
62	C00114	Choline	C5H14NO	104.107539
63	C00588	Choline phosphate	C5H15NO4P	184.073869
64	C00251	Chorismate	C10H10O6	226.047738
65	C00158	Citrate	C6H8O7	192.027003
66	C00327	L-Citrulline	C6H13N3O3	175.095691
67	C00055	CMP	C9H14N3O8P	323.051851
68	C00011	CO2	CO2	43.989829
69	C00010	CoA	C21H36N7O16P3S	767.115208
70	C05768	Coproporphyrinogen I	C36H44N4O8	660.315914
71	C00063	CTP	C9H16N3O14P3	482.984512
72	C00575	3',5'-Cyclic AMP	C10H12N5O6P	329.05252
73	C00941	3',5'-Cyclic CMP	C9H12N3O7P	305.041286
74	C00968	3',5'-Cyclic dAMP	C10H12N5O5P	313.057605

@BCL@E00B0CC3.xls

#	KEGG ID	Metabolite	Formula	Monoisotopic mass
75	C00942	3',5'-Cyclic GMP	C10H12N5O7P	345.047434
76	C00943	3',5'-Cyclic IMP	C10H11N4O7P	330.036535
77	C01419	Cys-Gly	C5H10N2O3S	178.041213
78	C02291	L-Cystathionine	C7H14N2O4S	222.067428
79	C00097	L-Cysteine	C3H7NO2S	121.019749
80	C00475	Cytidine	C9H13N3O5	243.085521
81	C00380	Cytosine	C4H5N3O	111.043262
82	C00206	dADP	C10H15N5O9P2	411.0345
83	C00360	dAMP	C10H14N5O6P	331.06817
84	C00131	dATP	C10H16N5O12P3	491.000831
85	C00705	dCDP	C9H15N3O10P2	387.023267
86	C00239	dCMP	C9H14N3O7P	307.056936
87	C00458	dCTP	C9H16N3O13P3	466.989597
88	C00857	Deamino-NAD+	C21H27N6O15P2	665.100962
89	C04691	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate	C7H13O10P	288.024633
90	C00966	2-Dehydropantoate	C6H10O4	146.057909
91	C00944	3-Dehydroquinate	C7H10O6	190.047738
92	C02637	3-Dehydroshikimate	C7H8O5	172.037173
93	C02934	3-Dehydrosphinganine	C18H37NO2	299.282429
94	C03226	3-Demethylubiquinone-9	C52H78O4	766.590011
95	C00559	Deoxyadenosine	C10H13N5O3	251.101839
96	C00881	Deoxycytidine	C9H13N3O4	227.090606
97	C00673	2-Deoxy-D-ribose 5-phosphate	C5H11O7P	214.024239
98	C00330	Deoxyguanosine	C10H13N5O4	267.096754
99	C05512	Deoxyinosine	C10H12N4O4	252.085855
100	C01801	Deoxyribose	C5H10O4	134.057909
101	C03496	Deoxy-ribose 1-phosphate	C5H11O7P	214.024239
102	C00526	Deoxyuridine	C9H12N2O5	228.074622
103	C00882	Dephospho-CoA	C21H35N7O13P2S	687.148878
104	C01909	Dethiobiotin	C10H18N2O3	214.131742
105	C00361	dGDP	C10H15N5O10P2	427.029415
106	C00362	dGMP	C10H14N5O7P	347.063084
107	C00286	dGTP	C10H16N5O13P3	506.995745
108	C01304	2,5-Diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine	C9H16N5O8P	353.073649
109	C01037	7,8-Diaminononanoate	C9H20N2O2	188.152478
110	C00986	1,3-Diaminopropane	C3H10N2	74.084398
111	C00415	Dihydrofolate	C19H21N7O6	443.155331
112	C05925	Dihydroneopterin phosphate	C9H14N5O7P	335.063084
113	C00337	(S)-Dihydroorotate	C5H6N2O4	158.032757
114	C00921	Dihydropteroate	C14H14N6O3	314.112738
115	C04272	(R)-2,3-Dihydroxy-3-methylbutanoate	C5H10O4	134.057909
116	C04332	6,7-Dimethyl-8-(1-D-ribityl)lumazine	C13H18N4O6	326.122634
117	C00235	Dimethylallyl diphosphate	C5H12O7P2	246.005826
118	C01143	(R)-5-Diphosphomevalonate	C6H14O10P2	308.00622
119	C00363	dTDP	C10H16N2O11P2	402.022932
120	C00364	dTMP	C10H15N2O8P	322.056602
121	C00459	dTTP	C10H17N2O14P3	481.989263
122	C01346	dUDP	C9H14N2O11P2	388.007282
123	C00365	dUMP	C9H13N2O8P	308.040952
124	C00460	dUTP	C9H15N2O14P3	467.973613
125	C05400	Epimelibiose	C12H22O11	342.116212
126	C01054	(S)-2,3-Epoxysqualene	C30H50O	426.386166
127	C01694	Ergosterol	C28H44O	396.339216
128	C00279	D-Erythrose 4-phosphate	C4H9O7P	200.008589
129	C00469	Ethanol	C2H6O	46.041865
130	C00189	Ethanolamine	C2H7NO	61.052764
131	C00346	Ethanolamine phosphate	C2H8NO4P	141.019094
132	C00016	FAD	C27H33N9O15P2	785.157134
133	C01352	FADH2	C27H35N9O15P2	787.172785
134	C00448	trans,trans-Farnesyl diphosphate	C15H28O7P2	382.131026
135	C00125	Ferricytochrome c	C42H54FeN8O6S2	886.295715
136	C00126	Ferrocytochrome c	C42H54FeN8O6S2	886.295715
137	C00061	FMN	C17H21N4O9P	456.104615
138	C00067	Formaldehyde	CH2O	30.010565
139	C04640	2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine	C8H16N3O8P	313.067501
140	C00058	Formate	CH2O2	46.005479
141	C02700	L-Formylkynurenine	C11H12N2O4	236.079707
142	C00234	10-Formyltetrahydrofolate	C20H23N7O7	473.165896
143	C00095	D-Fructose	C6H12O6	180.063388
144	C05378	beta-D-Fructose 1,6-bisphosphate	C6H14O12P2	339.996049
145	C01094	D-Fructose 1-phosphate	C6H13O9P	260.029719
146	C00665	D-Fructose 2,6-bisphosphate	C6H14O12P2	339.996049
147	C05345	beta-D-Fructose 6-phosphate	C6H13O9P	260.029719
148	C02095	beta-D-Fucose	C6H12O5	164.068473

@BCL@E00B0CC3.xls

#	KEGG ID	Metabolite	Formula	Monoisotopic mass
149	C00122	Fumarate	C4H4O4	116.010959
150	C01061	4-Fumarylacetoacetate	C8H8O6	200.032088
151	C00124	D-Galactose	C6H12O6	180.063388
152	C03384	D-Galactose 1-phosphate	C6H13O9P	260.029719
153	C05401	Galactosylglycerol	C9H18O8	254.100168
154	C01235	1-alpha-D-Galactosyl-myo-inositol	C12H22O11	342.116212
155	C00035	GDP	C10H15N5O11P2	443.024329
156	C00096	GDPmannose	C16H25N5O16P2	605.077153
157	C00341	Geranyl diphosphate	C10H20O7P2	314.068426
158	C01722	L-Glucitol	C6H14O6	182.079038
159	C01236	D-Glucono-1,5-lactone 6-phosphate	C6H11O9P	258.014068
160	C00329	D-Glucosamine	C6H13NO5	179.079373
161	C06156	D-Glucosamine 1-phosphate	C6H14NO8P	259.045703
162	C00352	D-Glucosamine 6-phosphate	C6H14NO8P	259.045703
163	C00267	alpha-D-Glucose	C6H12O6	180.063388
164	C00221	beta-D-Glucose	C6H12O6	180.063388
165	C00103	D-Glucose 1-phosphate	C6H13O9P	260.029719
166	C00668	alpha-D-Glucose 6-phosphate	C6H13O9P	260.029719
167	C01172	beta-D-Glucose 6-phosphate	C6H13O9P	260.029719
168	C00025	L-Glutamate	C5H9NO4	147.053158
169	C01165	L-Glutamate 5-semialdehyde	C5H9NO3	131.058243
170	C00064	L-Glutamine	C5H10N2O3	146.069142
171	C03734	alpha-D-Glutamyl phosphate	C5H10NO7P	227.019488
172	C00669	gamma-L-Glutamyl-L-cysteine	C8H14N2O5S	250.062342
173	C00051	Glutathione	C10H17N3O6S	307.083806
174	C00118	D-Glyceraldehyde 3-phosphate	C3H7O6P	169.998024
175	C00116	Glycerol	C3H8O3	92.047344
176	C00093	sn-Glycerol 3-phosphate	C3H9O6P	172.013675
177	C00184	Glycerone	C3H6O3	90.031694
178	C00111	Glycerone phosphate	C3H7O6P	169.998024
179	C00037	Glycine	C2H5NO2	75.032028
180	C00182	Glycogen	C24H42O21	666.221858
181	C00266	Glycolaldehyde	C2H4O2	60.021129
182	C00048	Glyoxylate	C2H2O3	74.000394
183	C00144	GMP	C10H14N5O8P	363.057999
184	C00044	GTP	C10H16N5O14P3	522.99066
185	C03078	4-Guanidinobutanamide	C5H12N4O	144.101111
186	C01035	4-Guanidinobutanoate	C5H11N3O2	145.085127
187	C00242	Guanine	C5H5N5O	151.04941
188	C00387	Guanosine	C10H13N5O5	283.091669
189	C00027	H2O2	H2O2	34.005479
190	C00032	Heme	C34H34FeN4O4	618.192948
191	C00135	L-Histidine	C6H9N3O2	155.069477
192	C00860	L-Histidinol	C6H11N3O	141.090212
193	C01100	L-Histidinol phosphate	C6H12N3O4P	221.056542
194	C05330	Homocysteine	C4H9NO2S	135.035399
195	C00544	Homogentisate	C8H8O4	168.042259
196	C05662	Homoisocitrate	C7H10O7	206.042653
197	C00263	L-Homoserine	C4H9NO3	119.058243
198	C00283	Hydrogen sulfide	H2S	33.987721
199	C06237	(R)-3-Hydroxy-3-methyl-2-oxobutanoate	C5H8O4	132.042259
200	C00356	(S)-3-Hydroxy-3-methylglutaryl-CoA	C27H44N7O20P3S	911.157467
201	C00632	3-Hydroxyanthranilate	C7H7NO3	153.042593
202	C00156	4-Hydroxybenzoate	C7H6O3	138.031694
203	C01251	2-Hydroxybutane-1,2,4-tricarboxylate	C7H10O7	206.042653
204	C04294	5-(2-Hydroxyethyl)-4-methylthiazole	C6H9NOS	143.040485
205	C02794	3-Hydroxykynurenine	C10H12N2O4	224.079707
206	C01157	trans-4-Hydroxy-L-proline	C5H9NO3	131.058243
207	C06056	4-Hydroxy-L-threonine	C4H9NO4	135.053158
208	C01024	Hydroxymethylbilane	C40H46N4O17	854.285796
209	C01179	3-(4-Hydroxyphenyl)pyruvate	C9H8O4	180.042259
210	C00262	Hypoxanthine	C5H4N4O	136.038511
211	C00104	IDP	C10H14N4O11P2	428.01343
212	C01267	3-(Imidazol-4-yl)-2-oxopropyl phosphate	C6H9N2O5P	220.024908
213	C04666	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate	C6H11N2O6P	238.035473
214	C00130	IMP	C10H13N4O8P	348.0471
215	C02693	Indole-3-acetamide	C10H10N2O	174.079313
216	C00954	Indole-3-acetate	C10H9NO2	175.063329
217	C00954	Indoleacetate	C10H9NO2	175.063329
218	C02938	3-Indoleacetonitrile	C10H8N2	156.068748
219	C03506	Indoleglycerol phosphate	C11H14NO6P	287.055874
220	C00294	Inosine	C10H12N4O5	268.08077
221	C00137	myo-Inositol	C6H12O6	180.063388
222	C01245	D-myo-Inositol 1,4,5-trisphosphate	C6H15O15P3	419.962379

#	KEGG ID	Metabolite	Formula	Monoisotopic mass
223	C01177	1L-myo-Inositol 1-phosphate	C6H13O9P	260.029719
224	C00311	Isocitrate	C6H8O7	192.027003
225	C00407	L-Isoleucine	C6H13NO2	131.094629
226	C00129	Isopentenyl diphosphate	C5H12O7P2	246.005826
227	C02504	2-Isopropylmalate	C7H12O5	176.068473
228	C04411	3-Isopropylmalate	C7H12O5	176.068473
229	C02631	2-Isopropylmaleate	C7H10O4	158.057909
230	C00490	Itaconate	C5H6O4	130.026609
231	C00531	Itaconyl-CoA	C26H40N7O19P3S	879.131252
232	C00081	ITP	C10H15N4O14P3	507.979761
233	C00328	L-Kynurenine	C10H12N2O3	208.084792
234	C00424	(S)-Lactaldehyde	C3H6O2	74.036779
235	C00256	(R)-Lactate	C3H6O3	90.031694
236	C00186	(S)-Lactate	C3H6O3	90.031694
237	C03451	(R)-S-Lactoylglutathione	C13H21N3O8S	379.104935
238	C01724	Lanosterol	C30H50O	426.386166
239	C00123	L-Leucine	C6H13NO2	131.094629
240	C00248	Lipoamide	C8H15NOS2	205.059505
241	C00047	L-Lysine	C6H14N2O2	146.105528
242	C00711	Malate	C4H6O5	134.021523
243	C00383	Malonate	C3H4O4	104.010959
244	C00083	Malonyl-CoA	C24H38N7O19P3S	853.115602
245	C00208	Maltose	C12H22O11	342.116212
246	C00392	D-Mannitol	C6H14O6	182.079038
247	C00644	D-Mannitol 1-phosphate	C6H15O9P	262.045369
248	C00936	alpha-D-Mannose	C6H12O6	180.063388
249	C03812	alpha-D-Mannose 1-phosphate	C6H13O9P	260.029719
250	C00275	D-Mannose 6-phosphate	C6H13O9P	260.029719
251	C05399	Melibiitol	C12H24O11	344.131862
252	C05402	Melibiose	C12H22O11	342.116212
253	C01438	Methane	CH4	16.0313
254	C00409	Methanethiol	CH4S	48.003371
255	C00445	5,10-Methenyltetrahydrofolate	C20H22N7O6	456.163156
256	C00073	L-Methionine	C5H11NO2S	149.051049
257	C00141	3-Methyl-2-oxobutanoate	C5H8O3	116.047344
258	C04752	2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate	C6H11N3O7P2	299.007223
259	C04327	4-Methyl-5-(2-phosphoethyl)-thiazole	C6H10NO4PS	223.006815
260	C00143	5,10-Methylenetetrahydrofolate	C20H23N7O6	457.170982
261	C00546	Methylglyoxal	C3H4O2	72.021129
262	C03460	2-Methylprop-2-enoyl-CoA	C25H40N7O17P3S	835.141423
263	C00440	5-Methyltetrahydrofolate	C20H25N7O6	459.186632
264	C04489	5-Methyltetrahydropteroyltri-L-glutamate	C25H36N8O12	640.245269
265	C00170	5'-Methylthioadenosine	C11H15N5O3S	297.08956
266	C05103	4alpha-Methylzymosterol	C28H46O	398.354866
267	C00418	(R)-Mevalonate	C6H12O4	148.073559
268	C06424	Myristic acid	C14H28O2	228.20893
269	C04302	N-(5-Phospho-D-ribosyl)anthranilate	C12H16NO9P	349.056268
270	C03406	N-(L-Arginino)succinate	C10H18N4O6	290.122634
271	C01152	N(pai)-Methyl-L-histidine	C7H11N3O2	169.085127
272	C03883	N1-(5-Phospho-D-ribosyl)-AMP	C15H25N5O14P2	561.087324
273	C00612	N1-Acetylspermidine	C9H21N3O	187.168462
274	C02567	N1-Acetylspermine	C12H28N4O	244.226312
275	C00437	N2-Acetyl-L-ornithine	C7H14N2O3	174.100442
276	C03794	N6-(1,2-Dicarboxyethyl)-AMP	C14H18N5O11P	463.074043
277	C00449	N6-(L-1,3-Dicarboxypropyl)-L-lysine	C11H20N2O6	276.132136
278	C00140	N-Acetyl-D-glucosamine	C8H15NO6	221.089937
279	C04256	N-Acetyl-D-glucosamine 1-phosphate	C8H16NO9P	301.056268
280	C00357	N-Acetyl-D-glucosamine 6-phosphate	C8H16NO9P	301.056268
281	C00624	N-Acetyl-L-glutamate	C7H11NO5	189.063722
282	C04133	N-Acetyl-L-glutamate 5-phosphate	C7H12NO8P	269.030053
283	C01250	N-Acetyl-L-glutamate 5-semialdehyde	C7H11NO4	173.068808
284	C02714	N-Acetylputrescine	C6H14N2O	130.110613
285	C00003	NAD+	C21H28N7O14P2	664.116947
286	C00004	NADH	C21H29N7O14P2	665.124772
287	C00006	NADP+	C21H29N7O17P3	744.083277
288	C00005	NADPH	C21H30N7O17P3	745.091102
289	C00438	N-Carbamoyl-L-aspartate	C5H8N2O5	176.043321
290	C00014	NH3	H3N	17.026549
291	C00153	Nicotinamide	C6H6N2O	122.048013
292	C00253	Nicotinate	C6H5NO2	123.032028
293	C01185	Nicotinate D-ribonucleotide	C11H15NO9P	336.048443
294	C04145	all-trans-Nonaprenyl diphosphate	C45H76O7P2	790.506628
295	C02571	O-Acetylcarnitine	C9H18NO4	204.123583
296	C01077	O-Acetyl-L-homoserine	C6H11NO4	161.068808

@BCL@E00B0CC3.xls

#	KEGG ID	Metabolite	Formula	Monoisotopic mass
297	C00979	O-Acetyl-L-serine	C5H9NO4	147.053158
298	C06055	O-Phospho-4-hydroxy-L-threonine	C4H10NO7P	215.019488
299	C01102	O-Phospho-L-homoserine	C4H10NO6P	199.024574
300	C00077	L-Ornithine	C5H12N2O2	132.089878
301	C00295	Orotate	C5H4N2O4	156.017107
302	C01103	Orotidine 5'-phosphate	C10H13N2O11P	368.025696
303	C00009	Orthophosphate	H3O4P	97.976895
304	C01118	O-Succinyl-L-homoserine	C8H13NO6	219.074287
305	C00036	Oxaloacetate	C4H4O5	132.005873
306	C05533	Oxaloglutarate	C7H8O7	204.027003
307	C05379	Oxalosuccinate	C6H6O7	190.011353
308	C00127	Oxidized glutathione	C20H32N6O12S2	612.151962
309	C06054	2-Oxo-3-hydroxy-4-phosphobutanoate	C4H7O8P	213.987854
310	C00322	2-Oxoadipate	C6H8O5	160.037173
311	C00109	2-Oxobutanoate	C4H6O3	102.031694
312	C00026	2-Oxoglutarate	C5H6O5	146.021523
313	C00141	2-Oxoisovalerate	C5H8O3	116.047344
314	C00007	Oxygen	O2	31.989829
315	C01260	P1,P4-Bis(5'-adenosyl) tetraphosphate	C20H28N10O19P4	836.048265
316	C01261	P1,P4-Bis(5'-guanosyl) tetraphosphate	C20H28N10O21P4	868.038094
317	C00249	Palmitate	C16H32O2	256.24023
318	C00154	Palmitoyl-CoA	C37H66N7O17P3S	1005.344874
319	C01134	Pantetheine 4'-phosphate	C11H23N2O7PS	358.096358
320	C00522	(R)-Pantoate	C6H12O4	148.073559
321	C00864	(R)-Pantothenate	C9H17NO5	219.110673
322	C02505	2-Phenylacetamide	C8H9NO	135.068414
323	C07086	Phenylacetic acid	C8H8O2	136.05243
324	C00079	L-Phenylalanine	C9H11NO2	165.078979
325	C00166	Phenylpyruvate	C9H8O3	164.047344
326	C00053	3'-Phosphoadenylyl sulfate	C10H15N5O13P2S	506.986229
327	C00119	5-Phospho-alpha-D-ribose 1-diphosphate	C5H13O14P3	389.951815
328	C00345	6-Phospho-D-gluconate	C6H13O10P	276.024633
329	C00631	2-Phospho-D-glycerate	C3H7O7P	185.992939
330	C00197	3-Phospho-D-glycerate	C3H7O7P	185.992939
331	C00236	3-Phospho-D-glyceroyl phosphate	C3H8O10P2	265.959269
332	C04751	1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate	C9H14N3O9P	339.046766
333	C04896	5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide	C15H25N5O15P2	577.082238
334	C00074	Phosphoenolpyruvate	C3H5O6P	167.982374
335	C03082	4-Phospho-L-aspartate	C4H8NO7P	213.003838
336	C01107	(R)-5-Phosphomevalonate	C6H13O7P	228.039889
337	C03232	3-Phosphonooxypyruvate	C3H5O7P	183.977289
338	C03492	D-4'-Phosphopantothenate	C9H18NO8P	299.077003
339	C04352	(R)-4'-Phosphopantothenoyl-L-cysteine	C12H23N2O9PS	402.086188
340	C04823	1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole	C13H19N4O12P	454.073709
341	C04677	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	C9H15N4O8P	338.06275
342	C04734	1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	C10H15N4O9P	366.057665
343	C03090	5-Phosphoribosylamine	C5H12NO7P	229.035138
344	C03838	5'-Phosphoribosylglycinamide	C7H15N2O8P	286.056602
345	C04376	5'-Phosphoribosyl-N-formylglycinamide	C8H15N2O9P	314.051517
346	C01005	3-Phosphoserine	C3H8NO6P	185.008924
347	C12144	Phytosphingosine	C18H39NO3	317.292994
348	C02656	Pimelic acid	C7H12O4	160.073559
349	C00931	Porphobilinogen	C10H14N2O4	226.095357
350	C00254	Prephenate	C10H10O6	226.047738
351	C00148	L-Proline	C5H9NO2	115.063329
352	C02191	Protoporphyrin	C34H34N4O4	562.258006
353	C01079	Protoporphyrinogen IX	C34H40N4O4	568.304956
354	C01168	Pseudouridine 5'-phosphate	C9H13N2O9P	324.035867
355	C00134	Putrescine	C4H12N2	88.100048
356	C03722	Pyridine-2,3-dicarboxylate	C7H5NO4	167.021858
357	C00250	Pyridoxal	C8H9NO3	167.058243
358	C00018	Pyridoxal phosphate	C8H10NO6P	247.024574
359	C00534	Pyridoxamine	C8H12N2O2	168.089878
360	C00647	Pyridoxamine phosphate	C8H13N2O5P	248.056208
361	C00314	Pyridoxine	C8H11NO3	169.073893
362	C00627	Pyridoxine phosphate	C8H12NO6P	249.040224
363	C00013	Pyrophosphate	H4O7P2	177.943226
364	C04281	L-1-Pyrroline-3-hydroxy-5-carboxylate	C5H7NO3	129.042593
365	C03912	(S)-1-Pyrroline-5-carboxylate	C5H7NO2	113.047678
366	C00022	Pyruvate	C3H4O3	88.016044
367	C00492	Raffinose	C18H32O16	504.169035
368	C01684	D-Rhamnose	C6H12O5	164.068473
369	C00255	Riboflavin	C17H20N4O6	376.138284
370	C00121	D-Ribose	C5H10O5	150.052823

@BCL@E00B0CC3.xls

#	KEGG ID	Metabolite	Formula	Monoisotopic mass
371	C00442	alpha-D-Ribose 1-phosphate	C5H11O8P	230.019154
372	C00620	D-Ribose 1-phosphate	C5H11O8P	230.019154
373	C00117	D-Ribose 5-phosphate	C5H11O8P	230.019154
374	C00199	D-Ribulose 5-phosphate	C5H11O8P	230.019154
375	C04425	S-Adenosyl-4-methylthio-2-oxobutanoate	C15H20N5O6S	398.113429
376	C00021	S-Adenosyl-L-homocysteine	C14H20N6O5S	384.121588
377	C00019	S-Adenosyl-L-methionine	C15H23N6O5S	399.145064
378	C01137	S-Adenosylmethioninamine	C14H23N6O3S	355.155234
379	C00447	Sedoheptulose 1,7-bisphosphate	C7H16O13P2	370.006614
380	C00281	Sedoheptulose 7-phosphate	C7H15O10P	290.040283
381	C00065	L-Serine	C3H7NO3	105.042593
382	C01031	S-Formylglutathione	C11H17N3O7S	335.078721
383	C00493	Shikimate	C7H10O5	174.052823
384	C03175	Shikimate 3-phosphate	C7H11O8P	254.019154
385	C05778	Sirohydrochlorin	C42H46N4O16	862.290881
386	C03172	S-Methyl-L-methionine	C6H14NO2S	164.074524
387	C00794	D-Sorbitol	C6H14O6	182.079038
388	C01452	Sorbose	C6H12O6	180.063388
389	C00315	Spermidine	C7H19N3	145.157898
390	C00750	Spermine	C10H26N4	202.215747
391	C00836	Sphinganine	C18H39NO2	301.29808
392	C01120	Sphinganine 1-phosphate	C18H40NO5P	381.26441
393	C00751	Squalene	C30H50	410.391252
394	C01530	Stearate	C18H36O2	284.27153
395	C00042	Succinate	C4H6O4	118.026609
396	C00232	Succinate semialdehyde	C4H6O3	102.031694
397	C00091	Succinyl-CoA	C25H40N7O19P3S	867.131252
398	C00089	Sucrose	C12H22O11	342.116212
399	C00059	Sulfate	H2O4S	97.967379
400	C00094	Sulfite	H2O3S	81.972465
401	C03785	D-Tagatose 1,6-bisphosphate	C6H14O12P2	339.996049
402	C01097	D-Tagatose 6-phosphate	C6H13O9P	260.029719
403	C02593	Tetradecanoyl-CoA	C35H62N7O17P3S	977.313574
404	C00101	Tetrahydrofolate	C19H23N7O6	445.170982
405	C04144	Tetrahydropteroyltri-L-glutamate	C24H34N8O12	626.229619
406	C00378	Thiamin	C12H17N4OS	265.112307
407	C00068	Thiamin diphosphate	C12H19N4O7P2S	425.044968
408	C01081	Thiamin monophosphate	C12H18N4O4PS	345.078637
409	C03028	Thiamin triphosphate	C12H20N4O10P3S	505.011298
410	C00188	L-Threonine	C4H9NO3	119.058243
411	C00214	Thymidine	C10H14N2O5	242.090272
412	C00178	Thymine	C5H6N2O2	126.042927
413	C01083	alpha,alpha-Trehalose	C12H22O11	342.116212
414	C00689	alpha,alpha'-Trehalose 6-phosphate	C12H23O14P	422.082542
415	C00078	L-Tryptophan	C11H12N2O2	204.089878
416	C00082	L-Tyrosine	C9H11NO3	181.073893
417	C01967	Ubiquinone-9	C54H82O4	794.621311
418	C00015	UDP	C9H14N2O12P2	404.002197
419	C00052	UDP-D-galactose	C15H24N2O17P2	566.05502
420	C00029	UDPglucose	C15H24N2O17P2	566.05502
421	C00203	UDP-N-acetyl-D-galactosamine	C17H27N3O17P2	607.081569
422	C00105	UMP	C9H13N2O9P	324.035867
423	C03543	Undecaprenyl diphosphate	C55H92O7P2	926.631828
424	C00106	Uracil	C4H4N2O2	112.027277
425	C00086	Urea	CH4N2O	60.032363
426	C01010	Urea-1-carboxylate	C2H4N2O3	104.022192
427	C00603	(S)-Ureidoglycolate	C3H6N2O4	134.032757
428	C00299	Uridine	C9H12N2O6	244.069536
429	C01051	Uroporphyrinogen III	C40H44N4O16	836.275231
430	C00075	UTP	C9H15N2O15P3	483.968527
431	C00183	L-Valine	C5H11NO2	117.078979
432	C00385	Xanthine	C5H4N4O2	152.033425
433	C01762	Xanthosine	C10H12N4O6	284.075684
434	C00655	Xanthosine 5'-phosphate	C10H13N4O9P	364.042015
435	C00181	D-Xylose	C5H10O5	150.052823
436	C06814	D-Xylose-5-phosphate	C5H11O8P	230.019154
437	C00310	D-Xylulose	C5H10O5	150.052823
438	C05437	Zymosterol	C27H44O	384.339216

ESM Table S3

Click here to download Table: S3_pos_ESI_tandem-MS.xls

@BCL@E00B15CC.xls

ID	Metabolite	Morther ion	Daughter ion	Loss
00001	4-Aminobutyric acid	104.07	87.04	NH3
			86.06	H2O
			69.03	NH3 + H2O
00003	L-Alanine	90.06	62.06	CO
00004	L-Arginine	175.14	116.08	CH5N3
			60.06	C5H9NO2
			158.11	NH3
			70.07	CH5N3 + CHOOH
			130.11	CO + NH3
			157.13	H2O
			72.08	C4H9NO2
			112.1	CHOOH + NH3
00005	L-Aspartic acid	134.06	88.04	CHOOH
			74.02	CH3COOH
			116.04	H2O
00006	L-Asparagine	133.0766	87.06	CHOOH
			116.04	NH3
			74.02	CH3CONH2
			88.04	CHONH2
00008	L-Cysteine	122.04	76.02	CHOOH
			105.01	NH3
			86.99	NH3 + H2O
00009	L-Glutamic acid	148.08	130.06	H2O
			102.06	CHOOH
			84.04	CHOOH + H2O
00010	L-Glutamine	147.09	130.06	NH3
			84.04	CHONH2 + H2O
			101.08	CHOOH
00011	L-Histidine	156.0958	110.08	CHOOH
			95.06	
00012	L-Homocystine	269.09	136.06	C4H7NO2S
			134.04	C4H9NO2S
			118.04	C4H7NO2S + H2O
00014	L-Isoleucine	132.12	86.1	CHOOH
			58.066	
00015	L-Leucine	132.12	86.1	CHOOH
			69.07	N?
00016	L-Lysine	147.13	130.1	NH3
			84.085	CHOOH + NH3
00017	L-Methionine	150.0775	104.0614	CHOOH
			133.05	NH3
			102.0634	CH4S
			56.05	CH4S + CO + H2O
			61.01	
00019	L-Phenylalanine	166.1	120.094	CHOOH
			149.08	NH3
			131.0636	NH3 + H2O
00021	L-Serine	106.0618	60.04	CHOOH
			88.04	H2O
			70.03	2H2O

@BCL@E00B15CC.xls

ID	Metabolite	Morther ion	Daughter ion	Loss
00023	L-Tryptophan	205.1241	188.0925	NH3
			159.11	CHOOH
			146.08	
00024	L-Tyrosine	182.1039	165.07	NH3
			136.0941	CHOOH
			123.0598	
			147.063	NH3 + H2O
			119.0647	CHOOH + NH3
00025	L-Valine	118	72.0827	CHOOH
00027	O-acetyl-L-serine	148.06	88.04	CH2CO + H2O
			106.05	CH2CO
			60.05	CH2CO + CO + H2O
			70.03	CH2CO + 2H2O
			102.05	CHOOH
00028	O-Phospho-L-serine	186.04	88.0436	H3PO4
00029	N-Acetyl-L-glutamic acid	190.0951	130.07	CH2CO + H2O
			172.08	H2O
			148.08	CH2CO
			144.08	CHOOH
			102.06	CH2CO + CHOOH
			84.05	CH2CO + CHOOH + H2O
00030	Betaine	118.1003	59.07	CH3COOH
00033	Apidic acid	147.08	129.0698	H2O
			101.07	CHOOH
			111.06	2H2O
			83.05	CHOOH + H2O
00034	cis-Aconitic acid	175.03	139	2H2O
			157.01	H2O
			111.01	CHOOH + H2O
00035	Citraconic acid	131.05	113.037	H2O
			85.032	CHOOH
00036	Citramalic acid	149.06	103.05	CHOOH
			131.05	H2O
			85.03	CHOOH + H2O
00037	Citric acid	193.06	175.04	H2O
			157.02	2H2O
			129.03	CHOOH + H2O
			139.03	3 H2O
			147.05	CHOOH
00038	DL-Isocitric acid	193.06	129.03	CHOOH + H2O
			175.05	H2O
			157.04	2H2O
			147.05	CHOOH
00040	Glutaric acid	133.07	115.05	H2O
			87.05	CHOOH
00043	Itaconic acid	131.04	113.03	H2O
			85.03	CHOOH
00045	Malic acid	135.05	89.03	CHOOH
			117.033	H2O
			71.01	CHOOH + H2O

@BCL@E00B15CC.xls

ID	Metabolite	Morther ion	Daughter ion	Loss
00049	Pimelic acid	161.1	143.09	H2O
			125.08	2H2O
			97.07	CHOOH + H2O
			115.09	CHOOH
			69.07	2CHOOH
00051	α-Ketoapidic acid	161.05	143.09	H2O
00052	Flavin adenine dinucleotide (FAD)	786.16	348.07	
00054	NAD	664.1	542.07	
			524.03	
			428.03	C9H10N5O3
00057	AMP	348.07	136.06	
			213.01	
			330.06	H2O
00059	GTP	524.01	152.06	
00061	cGMP	346.06	152.06	
00063	Cytosine	112.0627	95.03	NH3
			69.05	
			94.05	H2O
00065	Acetyl-DL-carnitine	204.12	85.03	
			145.05	
			60.09	
00066	Fructose 1,6-bisphosphate	440.11	100.12	
			323.02	
			341.03	
00067	Phospho(enol)pyruvate	168.99	150.98	H2O
			122.99	CHOOH
			80.98	
00069	2-phosphoglyceric acid	0	98.99	
			140.99	
00071	Glucose 1-phosphate	261.05	98.98	C6H10O5
			163.06	H3PO4
00076	Trehalose	343.17	325.15	H2O
			163.09	C6H12O6
			181.0961	C6H10O5
00079	N-acetyl-D-glucosamine	222.1	204.09	H2O
			138.06	
			186.08	2H2O
			168.07	3 H2O
			126.07	CH3COOH + 2H2O
00081	Riboflavin	377.1902	243.11	C5H10O4

ESM Table S4

Click here to download Table: S4_neg_ESI_tandem-MS.xls

@BCL@E00BDFD4.xls

ID	Metabolite	Morther ion	Daughter ion	Loss
00001	4-Aminobutyric acid	102.05	n.d.	
00002	Glycine	74.03	n.d.	
00003	L-Alanine	88.04	n.d.	
00005	L-Aspartic acid	132.03	88.03	CO2
			115	NH3
00006	L-Asparagine	131.04	114.01	NH3
			70.04	NH3 + CO2
			113.04	H2O
00008	L-Cysteine	120.01	n.d.	
00009	L-Glutamic acid	146.04	102.06	CO2
			128.04	H2O
00010	L-Glutamine	145.05	127.03	H2O
			109.04	2H2O
00011	L-Histidine	154.06	n.d.	
00012	L-Homocystine	267.05	n.d.	
00017	L-Methionine	148.04	n.d.	
00019	L-Phenylalanine	164.07	147.05	NH3
			103.05	NH3 + CO2
			72.01	C7H8
00020	L-Proline	114.05	n.d.	
00021	L-Serine	104.03	74.02	CH2O
00023	L-Tryptophan	203.08	n.d.	
00024	L-Tyrosine	180.06	n.d.	
00025	L-Valine	116.06	n.d.	
00027	O-acetyl-L-serine	146.04	n.d.	
00028	O-Phospho-L-serine	183.99	96.97	C3H5NO2
00029	N-Acetyl-L-glutamic acid	188.04	128.03	CH3COOH
			102.05	
			144.06	CHOCH3 or CO2
			59.02	C5H7NO3
			170.04	H2O
			100.04	CHOCH3 + CO2
			146.04	
00031	2-Hydroxyisobutyric acid	103.04	57.04	CO + H2O
00032	6-phosphogluconic acid	275	96.96	C6H10O6
			177.04	H3PO4
			256.98	H2O
			78.95	C6H12O7
00033	Apidic acid	145.04	101.06	CO2
			83.05	CO2+ H2O
00034	cis-Aconitic acid	173	85.03	2CO2
			129.01	CO2
			111	CO2+ H2O
00035	Citraconic acid	129.02	85.02	CO2
00036	Citramalic acid	147.02	87.01	CH3COOH
			85.02	CO2+ H2O
			129.02	H2O
			57.05	
			103.04	CO2
			59.02	2CO2
00037	Citric acid	191	157.11	

@BCL@E00BDFD4.xls

ID	Metabolite	Morther ion	Daughter ion	Loss
			115.07	
			143.09	
			129.09	
00038	DL-Isocitric acid	191	157.11	
			115.07	
			143.09	
			129.09	
00039	Fumaric acid	115	n.d.	
00040	Glutaric acid	131.03	87.04	CO2
00041	Glyceric acid	105.02	n.d.	
00042	Glyoxylic acid	73	n.d.	
00043	Itaconic acid	129.02	n.d.	
00044	Lactic acid	89.02	n.d.	
00045	Malic acid	133.01	n.d.	
00046	Myristic acid	227.19	n.d.	
00047	Oxalacetic acid	131	87.01	CO2
00048	Oxalic acid	89	n.d.	
00049	Pimelic acid	159.06	97.07	CO2+ H2O
			141.05	H2O
			115.07	CO2
00050	Tartaric acid	149	87.01	CO2+ H2O
			73	
			59.02	
			103	CHOOH
			130.97	H2O
00051	α-Ketoapidic acid	159.02	59.03	
			115.04	CO2
			87.05	
00052	Flavin adenine dinucleotide (FAD)	784.13	n.d.	
00054	NAD	662	540.04	
00055	ATP	505.99	426.02	
00057	AMP	346.04	n.d.	
00059	GTP	522	n.d.	
00061	cGMP	344.05	n.d.	
00064	Coenzyme A	766.1	n.d.	
00066	Fructose 1,6-bisphosphate	338.98	96.97	C6H11O8P
			241	H3PO4
00067	Phospho(enol)pyruvate	166.97	78.96	2CO2
00068	3-phosphoglyceric acid	184.98	96.97	2CO2
			78.96	2CO2+ H2O
00069	2-phosphoglyceric acid	184.98	96.97	2CO2
			78.96	2CO2+ H2O
00071	Glucose 1-phosphate	259.01	241.01	H2O
			96.96	C6H10O5
00072	Galactose 6-phosphate	259.02	n.d.	
00074	Uridine 5'-diphospho-glucose	565.03	n.d.	
00075	Uridine 5'-diphospho-galactose	565.03	n.d.	
00076	Trehalose	341.11	n.d.	
00079	N-acetyl-D-glucosamine	220.08	n.d.	